The Art of Benefits Prediction and the Statistical Science of Post-Implementation Assessment

in

Aviation Investment Analysis

Stephen B. Cohen
<stephen.cohen@faa.gov>

October 18, 2000
(Revised June 15, 2001)

Federal Aviation Administration
Investment Analysis & Operations Research (ASD-400)

Notice

This document was originally published under the title *General Guidelines for Conducting the Benefits Analysis Portion of an Investment Analysis.*

Notice Regarding Microsoft Word Text and Equation Alteration

Microsoft Word® documents have the disconcerting habit of changing fonts and formatting when sent as e-mail attachments, and perhaps under other circumstances. If you believe that your copy of this document has been corrupted, please contact Steve Cohen at stephen.cohen@faa.gov .

Equations in Word® sometimes do not print correctly. This usually can be attributed to the printer driver. As each printer and driver is different, a one-fits-all solution is not available. However, the following example of a "fix" for a Hewlett-Packard LaserJet 4si® may suggest a "fix" for your printing problems.

Procedure for Correcting MS Word® Equation Printing on an HP LaserJet 4si® Printer Using the HP LaserJet 4Si/4Si MX Printer Driver

- In the document, click *Tools, Options, Save* (tab), *Embed True Type Fonts, OK*.
- Click on the Windows® *START* button, select *Settings, Printers.*
- Right-click on the printer you intend to use, and select *Properties*.
- Click on the *Print Quality* tab and select *Raster* and *True Type as Graphics*.
- Click *OK* and then close the Printers window.

Your document should now print properly.

If you cannot get your printer to properly print the equations, you may obtain a paper copy of this document by sending a request to stephen.cohen@faa.gov .

ACKNOWLEDGMENTS

Grateful thanks to the following people whose help was invaluable in creating this document.

Duc Le, who took a hodge-podge of information on data bases and models from the author and combined this with information from FAA Web sites to produce the highly useful, structured tables in Appendices D and E.

Arturo Politano, who provided the information on Risk Analysis and from whom the author more-than-liberally copied. (Any errors or omissions, however, are the author's.) Also for directing the project that developed an initial list of tools[1] which was used in developing Appendix E.

Dan Citrenbaum, Nastaran Coleman, Lewis Fisher, and Arturo Politano, who (in no particular order) furnished many wonderful ideas, together and separately, that are incorporated herein.

Madelyn Harp, who prepared an initial list of tools[2] which was used in developing Appendix E.

Alice Harball, who started this project and did the initial research, of which the author is most grateful and made good use.

The EUROCONTROL CARE-*INTEGRA* Project and its Technical Manager, Mr. Ian Wilson, which provided much of the material in Appendix A.

Fran Melone, who gave the author the opportunity and time to produce this document.

Stephen Cohen
September 8, 2000

[1] Harp, Madelyn, "Preliminary Report 6: Exploiting Tools for Investment Analysis," Washington: SETA and Operations Research and Analysis Branch, Investment Analysis and Operations Research, FAA, Jan. 30, 1998.

[2] Ibid.

(this page intentionally blank)

CONTENTS

APPENDICES

TABLES and FIGURES

CONDUCTING A BENEFIT ANALYSIS

PREFACE

This document serves as a companion to the Volpe report *Cost, Benefit, and Risk Assessment Guidelines for R,E&D Investment Portfolio Development*.[3] While the Volpe report provides a traditional description of benefit and other analyses and includes detailed guidance of some of the steps in a benefit analysis, this document takes a less traditional approach. The author hopes that this document will help the reader inexperienced in benefit analysis gain some intuitive feeling of the process as well as some warning of pitfalls that may be encountered. Special note should be made of the inclusion of several statistical tests for post-implementation project benefit assessment.

If you obtained this document as part of a compressed file, the file should also include a copy of the Volpe report. It is suggested that after the reader has reviewed a portion of the present document, s/he then review the corresponding material in the Volpe report, which should help with the "mechanics" and details of benefit prediction. The present document purposely does not provide minute details about benefit prediction. However, because information is not readily available on post-implementation assessment of a project's real-world effectiveness, this document does provide detailed statistical methodology for this topic. Please note that the Volpe report uses the term "assessment" for pre-implementation benefit prediction, while the term "assessment" in this document is used only in the context of post-implementation evaluation.

And so we begin **…**

The Product Team (PT) will have identified categories of benefits that it expects its product to deliver. It also should have reviewed how its product fits into the National Airspace System (NAS) architecture. However, it is unlikely that the PT members are very familiar with developing supportable benefit estimates. Because it is important that the PT understand and assist the Investment Analysis Team (IAT) in the IA process, it is useful to have a step-by-step process for conducting the benefit analysis. This should also help in starting the benefit estimation process early. Also, some PTs in their Mission Needs or other documents claim almost the entire universe as benefits for their project. Even if these claims have some validity, time, personnel, and data limitations make it impossible to quantify, let alone monetize, many of these potential benefits. Guidance in selecting for evaluation doable benefit areas is therefore important.

As with any effort, there are rules to follow. These may be found in Appendix A. Although the number of rules may seem excessive, if the benefit analysis generally follows the steps described below, it is unlikely that any of these rules will be violated.

---

[3] Report No. WP-43-FA92F-99-1, Cambridge: Operations Assessment Division, DTS-59, Volpe National Transportation Systems Center, October 1998.

However, it is a good idea to frequently satisfy yourself that the analysis has not strayed beyond the bounds of the rules.

Documentation is an important part of the process, not only for historical records, but also to help clarify issues. By putting something on paper and then reviewing what was written, one often discovers "holes" and new insights. Full documentation is also needed so that future IAs will have access to information needed to develop their reference cases (which may include the impacts of your project). It also is needed for post-implementation assessment of the impacts of your project, which the General Accounting Office (GAO) has "requested" the FAA to do. ***The steps below that should be documented are prefaced with the underlined Greek letter delta,*** <u>D</u>.

Documentation that is inappropriate for formal reports (possibly because of its detail) should be retained as part of the project file. Both paper and <u>electronic</u> copies of the project file should be placed in a central repository. The IA project leader should also retain paper and electronic copies. Far too often electronic copies of documentation produced by contractors has been lost.

Also, EVERYONE runs into unexpected difficulties. You will too, so start early.

Finally, beware the word "baseline," *which seems to have several meanings.*

The guidance is presented as a sequence of 14 steps with supporting material in appendices.

A. <u>FIRST STEPS ¾ THE PROJECT AND ITS POSSIBILITIES</u>

1. <u>D</u> Describe the project, including what and how it will "physically" and operationally change the NAS.

   For example, for ASDE-X, describe what it consists of and how it works: That is, include things like, "ASDE-X will locate and identify every aircraft on a runway or on a taxiway near a runway within ___ feet of its true position." At this stage, do not include statements like "ASDE-X will reduce runway accidents." Statements like the latter will come later.

<u>Do the following for each benefit area in which your project will have an impact.</u>

2. Identify the benefit category and its location in the "benefit universe." [The "benefit universe" is residing in Appendix B.]

   a) Use the diagrams to clarify where the benefits will accrue.

b) If there is uncertainty as to how to "locate" a category, ask yourself, "What will this project physically and operationally do? That is, in each diagram, decide which lowest-level box will be affected by the project and how it will be affected.

3. D Write a general description of what the future will be if your project is approved, proceeds as planned, and is successful.

a) Because the time value of money (net present value = NPV) is accounted for in the benefit analysis and because the system is forecast to change over time, you will later need to include year-by-year benefit estimates. So be sure to include in your general description any important dates, way points, etc. and what is significant about them.

4. D Write a description of the "reference case"…what is expected to occur if this project is not accomplished. (Later, you will monetize this scenario.)

a) There may be more than one possibility for a reference case. For example, if your project is to replace all equipment **X** at centers, the reference case might be to regularly perform maintenance on the existing **X** at each center or it may be to perform no maintenance until an **X** fails.

b) When there is more than one possibility, you can try to get an up-front decision from management as to which reference case to use, but you may have to determine (as described below) the impact of each possibility, before management will make a choice. You might even have to do a benefit analysis that presents (net present value) results using each possible reference case, if management does not make a choice.

c) Here, too, you should include any important dates, way points, etc. and what is significant about them.

B. PLANNING THE ANALYSIS

5. The Product Team (PT) will have determined how the project fits into the NAS Architecture, but it is important for you to check this as well. Visit the Architecture home page at http://www.nas-architecture.faa.gov. This page has links to several pages including the must-see *Capability Architecture Tool Suite* (CATS). Note that the version of CATS accessible from the home page may be different from the private FAA page, http://172.27.164.125/cats/

a) Ask yourself

i) On what does this project depend?

ii) What depends on this project?

iii) What other interactions are possible?

See Steps 9 and 10 for further guidance.

b) The Architecture is in a continual state of flux, so it is wise occasionally to check CATS for changes.

c) Other documents you may wish to check include

i) the *NAS Architecture Version 4 Report* (http://172.27.164.125/CATS/Tutorials/NASArch.htm)

ii) The *NAS Blueprint* (http://172.27.164.125/CATS/Tutorials/Blueprint.htm)

iii) The *FAA National Aviation Research Plan* (formerly the *RE&D Plan*) (http://172.27.164.125/CATS/Tutorials/NARP.htm)

iv) *Aviation Glossary* (http://172.27.164.125/CATS/Search/default.cfm?SG=TRUE)

v) Other related documents (http://172.27.164.125/CATS/Tutorials/Other-Intro.htm)

6. Discuss the anticipated benefit categories with individuals from the PT or, if necessary, elsewhere, who **directly** work in the areas that the project will impact.

a) Whenever possible, get your information from people who actually do the job(s) that might be impacted by the project. If possible and relevant, also watch them doing the job.

b) If you cannot get access to someone who actually does the job that might be impacted by the project, and instead you must obtain information from others, try to verify the information with additional sources.

c) It is surprising how often the way an "expert" insists things work is not the way they actually work.

d) **Ask probing questions.**

e) Try to arrange for an as-needed availability of your subject area experts.

f) You may need management assistance to obtain access to the expertise you need.

7. D Develop a plan for how the benefit estimation will be done.

a) **Benefits are usually first calculated as (changes in) metric values** such as reduced delays or fewer passengers killed. See Appendix C. Later these metric values are monetized (valued in dollars) to derive the final benefits values.

b) Because the time value of money is included in the benefits computations, benefits (changes in metric values) are usually computed on a yearly basis.

c) The answers to the following questions will help determine how the benefit analysis can be done and the depth to which it can or should be done.

i) What data are needed?

ii) What data are available?

- Data sources include ASD-400's PMAC, the Safety Office's NASDAC, NTSB's full Accident Reports and studies, Airway Facilities' NAPRS data base, and others. See Appendix D for more information on data sources

- Obtain up front solid commitments for access to the data you will need. In some cases you may have difficulty in getting an organization to share its data. Be aware of the possibility of stalling tactics. You may have to ask your management to intervene on your behalf. It is unwise to begin an IA without data access commitment.

- WARNING: Data are quite often other than what people (even you) believe them to represent. This is particularly true of coded data (as opposed to narratives). It is vital that you discuss the data with people who are intimately familiar with the data, preferably including both people who collect and people who use the data regularly.

iii) Are there relevant models or references available that can assist with the analysis and computations? See Appendices E and G.

iv) How much time is available? [A good rule of thumb is to plan on using only 2/3 of the time available so that you have time for unanticipated problems.]

v) What people resources are available to do the work?

vi) What funding resources are available?

vii) What is the project "visibility," and perceived or anticipated impact and value? If these are low, a quick-and-(not too) dirty analysis may be all that is called for. If these are high, fight for the resources needed to do a good job.

d) If the project has potential benefits in several areas, choose to work on and complete first those areas that are likely to show the greatest monetary benefit and for which you can compute the monetary benefits without too much difficulty.

e) In general, do not spend time on benefits that cannot be monetized, or for which monetization would be difficult. These benefits can be described qualitatively in your reports, but you probably will not have the resources to quantify them.

C. <u>ESTIMATING THE BENEFITS</u>

Note: Parts 8, 9, and 10 should be reviewed before beginning the benefit computation effort.

8. <u>D</u> Estimate the benefits.

a) Use the plan, data, and models you identified above.

b) As stated above, benefits are usually first calculated in terms of yearly changes in a metric, which are later monetized. The changes are computed as a difference in metric values:

(metric value with project in place) - (metric value in reference case situation)

c) If questions arise, make use of the subject matter experts with whom you previously made consulting arrangements.

d) Depending on the difficulty of computing benefits (metrics) and the resources available (including time), you may wish to compute benefits year-by-year or on a less frequent basis. In the later case, you can estimate benefits for the intermediate years by curve fitting or (not necessarily linear) interpolation.

9. <u>D</u> Check for the possibility that the program may have unintended, adverse consequences, particularly in the safety area. (The PT should have done this before the IA began, but you may have had new insights or discovered new information since then. Also the architecture or its time frame may have changed.)

a) A separate System Safety Assessment is now required as part of the Investment Analysis. This task is required whether or not it is believed that your project will have any adverse safety impacts. If it is found that your project may have adverse consequences, the PT will have to develop mitigants to ensure that the project doesn't reduce safety. The costs of these mitigants must be included in the IA. The results of the Safety Assessment will be reviewed by the ASD-110 Safety Team, presently led by Scott VanBuren. The IA team must plan for the time it takes ASD-110 to complete this review and for the possibility that the review may find the Safety Assessment to be inadequate.

b) If there are possible non-safety disbenefits, they need to be estimated.

c) Subtract the disbenefits from the benefits. (If there is, say, only an estimated 20% probability of incurring disbenefits, you may wish only to subtract 20% of the possible disbenefits from the benefits, or you may wish to provide both benefit values with no disbenefits included and benefit values with the maximum disbenefits included.)

10. D Check for double counting of benefits and the impact of other programs on your program's benefits.

a) Projects frequently are delayed. substantially modified, or even cancelled. It is therefore important to consider such impacts and their consequent effects on user and FAA benefits and costs. It must be remembered that benefits depend on time, not just in NPV sense, but also because delays may result in new technology so that a project may be overtaken by events.

b) It sometimes happens that another Investment Analysis has claimed benefits that your project is claiming. For example, if another project will serve as infrastructure for your project, the IA for that project may have claimed some of the benefits that actually will accrue only after your project becomes operational.

i) Only claim benefits that will directly accrue from the implementation of your project. If another project that will serve as infrastructure for yours has improperly claimed benefits that will only directly accrue from your project, then claim these benefits for your project, but also include in your report the information that the other project has claimed some of these benefits.

ii) A more sophisticated approach than this may be needed depending on the circumstances of the other project. For example,

- If the other project will only serve as infrastructure for your project alone, and it will produce no benefits other than those that would accrue as a result of your project's implementation, and the other project has not yet incurred any development or implementation expenses, then the IA Cost Team should include the costs of both projects and these costs should be compared with the benefits that would accrue from the implementation of both.

- If the other project will only serve as infrastructure for your project alone, and it will produce no benefits other than those that would accrue as a result of your project's implementation, and the other project has already been implemented, then its development and capital costs are "sunk" (already spent), and the IA Cost Team should include only its ongoing costs as part of the costs of achieving the benefits of your project.

- Most likely, the other project will serve as infrastructure for several projects. In this case, allocation of its costs against the benefits of these several projects can become quite complex and politics almost certainly will enter into the determination. Serious discussions with management are appropriate.

iii) It is also possible that another project may impact your reference case scenario in such a way as to reduce the size of the "problem" that your project would help mitigate. Figure 1 provides an illustration of this.

In this example, Project A will reduce equipment outages at TRACONs, thereby reducing outage-induced delays. Project B will develop better information on wake turbulence, thereby enabling closer arrival spacing of aircraft. This will result in a capacity increase at some busy airports. At busy airports, outage-induced delays can extend past the time the outage is ended because of the "stack-up" of aircraft. When Project B is operational, this "stack-up" will be reduced more quickly than would be the case if spacing were not reduced. Thus, Project B, when operational, has the effect of changing the reference case scenario of Project A, thereby reducing the benefits of Project A.

In this example, if the Investment Analysis for Project B preceded that of Project A, the benefit analysis for Project B would include estimates of the improved capacities at the affected TRACONs. The IA team for Project A could then use these estimated capacities to develop a new reference case for the years and locations where Project B is/will be operational. The benefit analysis of Project A would use this new reference case in estimating its benefits at the affected sites.

If it were uncertain whether Project B would become operational, then the Project A benefits analysis would include both benefit estimates assuming B would become operational and benefit estimates assuming B would not become operational.

If the IA for Project B did not have increased capacity estimates in time for use by Project A, then the benefits estimate for Project A would include both delay reduction estimates assuming Project B did not exist and at least qualitative estimates of the impact of an operational Project B on the benefits of Project A.

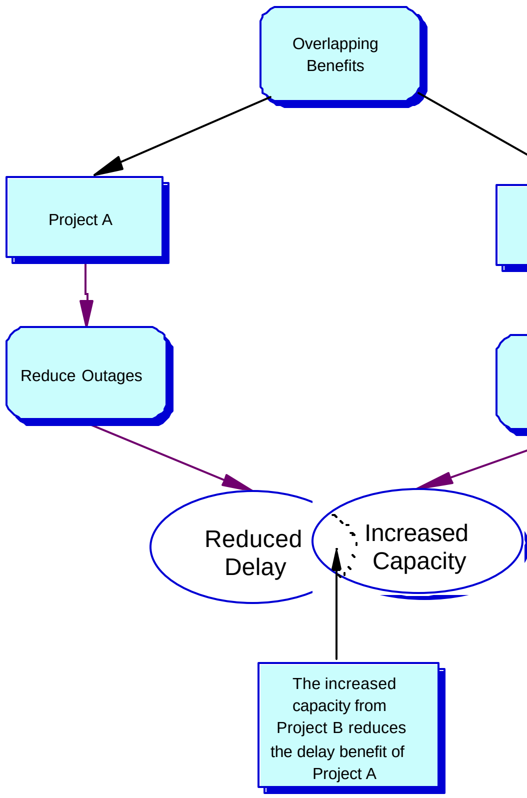

FIGURE 1
Reduced Benefits

11. D Net Present Value (NPV) benefit computation

a) If you haven't already done so, convert the yearly metric difference (project less reference case) values into monetary values using standard, official FAA, DOT and Federal values, such as in

i) *Economic Values for Evaluation of Federal Aviation Administration Investment and Regulatory Programs*, FAA-APO-98-8, June 1998, (or later). The latest version (as of May 2000) of this guide, which includes **an additional chapter not present in the paper version**, may be found at http://api.hq.faa.gov/apo_pubs.htm#ANCHOR98_10

Because this document may become obsolete, one should contact The FAA APO organization for current guidance. At present, we suggest contacting Stefan Hoffer (202- 267-3309) at APO.

ii) Other useful publications, data bases, and information may be found at http://api.hq.faa.gov/apo_pubs.htm and at http://www.apo.data.faa.gov/

b) Compute the net present value (NPV) of the benefits using the standard methodology and the current, official FAA and Office of Management and Budget (OMB) discount rate(s).

c) Unfortunately, the official document for discount rates, OMB Circular A-94, http://www.whitehouse.gov/OMB/circulars/a094/a094.html , does not present sufficient, clear guidance. It therefore is recommended that one use APO guidance provided in 11.a.i, above. Other APO documents may be useful. For a listing of these, go to http://api.hq.faa.gov/apo_pubs.htm . For a list of OMB guidance circulars, consult http://www.whitehouse.gov/OMB/circulars/index.html .

12. Risk analysis.

The risk analysis related to benefits should be an independent effort. However, the Risk Analysis Team will require documentation on the data and methodology used by the Benefits Team and will need to have access to members of the Benefits Team, Cost Team, Safety Assessment Team, and the PT. **It is, therefore, important that care be taken in maintaining the data used in the benefit analyses and in adequately documenting the methodologies and assumptions used.** Any concerns and/or uncertainties that surfaced during the benefit analysis should also be documented. Failure to maintain information required by the Risk Analysis Team may delay the completion of the Investment Analysis. The information below is provided to assist the Benefits Team in preparing the material needed for the Risk Analysis portion of the Investment Analysis.

Among the areas that the Risk Analysis Team will evaluate are the following:

a) Benefit Identification

i) Are the same benefits claimed by other programs? (Is there double counting?)

ii) Has a major benefit area been omitted?

iii) Are some of the benefits attributed to the program unrealistic? (Will the program REALLY be able to deliver them?)

> Are the benefits dependent on the existence of factors, such as other, non-completed programs, that may not be present at the time the benefits are supposed to be realized?

b) Benefit estimation

i) What assumptions were used in the benefit estimation and are they justified?

ii) How sensitive are the benefit estimates to changes in the assumptions?

iii) How reliable and appropriate are the data that were used.

iv) Were the benefit estimation techniques used appropriate and adequate, and did they account for all major factors needed to achieve the benefits?

v) Is the benefit analysis straightforward or tortuous?

vi) Were all calculations, including NPV calculations, done correctly, using standard FAA, DOT, and OMB values?

vii) Are the qualitative descriptions of non-quantifiable benefits reasonable.

viii) Are any estimates of cost avoidance reasonable, justifiable, and thorough. (Have all new expenses required to achieve the cost avoidance been included?)

c) The risk that the project may have unintended, adverse consequences.

The report, *Risk Assessment Guidelines for the Investment Analysis Process* is a good source of information. Other documents that contain information on risk are *Federal Aviation Administration Acquisition Management System* and *Cost, Benefit, and Risk Assessment Guidelines for R,E&D Investment Portfolio Development*.

D. POST-IMPLEMENTATION BENEFIT ASSESSMENT

Once a project has become operational, someone (the GAO, a Senator, or possibly the FAA itself) may be interested in assessing its impact: Has it improved the situation? Has it achieved the benefits that were claimed for it? (Did the FAA play "fast and loose" with the benefit estimates?)

Some Federal agencies, such as the DOT National Highway Traffic Safety Administration (NHTSA) and the recreational boating division of the U.S. Coast Guard, have been performing formal, post-implementation benefit assessments for over 20 years. The FAA, however, has seldom performed such assessments. The General Accounting Office (GAO) has suggested that the FAA perform such assessments, and at the time of this writing, the FAA is developing a formal process for doing so.

Irrespective of the specifics of any formalized process, however, the essence of a post-implementation assessment of the benefits of a project is the use of appropriate metrics and statistical methods.

Because of the existence of numerous sources of detailed information on benefits prediction, earlier parts of this document did not dwell on the minutiae of benefits analysis. However, the methodology for the post-implementation assessment of benefits tends to be in journal articles and in-house reports. As the FAA has seldom performed such analyses, it does not have a readily available source of detailed information on statistical techniques for post-implementation benefit assessment. For this reason, the remainder of this document will provide detailed guidance on such techniques.

13. Use metrics.

    Benefit assessment is normally performed using the same metric(s) that were used in the original IA benefit predictions. For instance, in the example of Figure C-2 in Appendix C, you may be interested in assessing how much outages have been reduced, how much capacity has been improved, or how much delay has been improved over what it would have been without the project(s). If you are interested in the monetary impact of the program, you should first calculate the benefit using metrics and then convert the results to monetary units, using both the monetary "constants" (e.g. passenger value of time, value of a life) that were used in the IA and using the values in existence during the period being assessed.

14. Use statistical methods

    Because real data is always subject to some random variation, statistical methods must be used in post-implementation benefit assessment. This is to ensure that any appearance of an impact is not just the result of the normal random variation in metric values that occur irrespective of any system change. For all but the simplest and least sensitive evaluation techniques, the services of a knowledgeable statistician are required.

There are two primary approaches to performing a statistical evaluation of the post-implementation performance of a project. One technique involves comparing actual post-implementation metric values to the reference case metric values that the original benefit analysis projected would occur were the project not implemented. This approach has a major pitfall. The original benefit analysis may have estimated that the "without project" future reference case metric values would be much worse than would have actually occurred. This may have been the result of erroneous assumptions about equipment or operational capabilities in future environments, or it may have been intentional so as to make the project appear more beneficial than really was expected.

The second, approach involves comparing actual post-implementation metric values to some statistical extrapolation of pre-implementation metric values. This approach is preferable, unless there are very good reasons to believe that future reference case metric values would be significantly worse than could be extrapolated from past metric values.

Table 1 suggests statistical methods appropriate to the depth of evaluation desired, the conditions the metric must meet, and the data that is available. The first six tests, Test A – Test F, are variations on the second, preferable approach to post-implementation statistical assessment of benefits. The final test, Test X, is based on the first, pitfall-prone approach to benefit assessment. Step-by-step descriptions of these tests may be found in Appendix F.

If at all possible, have a good statistician perform the post-implementation benefit assessment.

CAVEAT

Events exogenous to the program being evaluated can result in seriously confounding the data being analyzed. If post-implementation data appear to make no sense, you should investigate the following possibilities.

1) The way your metric data has been collected, recorded, created, or processed has changed.
2) The introduction or discontinuance of, or change in, other FAA programs or operational procedures may have affected the results of your program.
3) Changes in airline, air cargo, or other aircraft operations may have affected the results of your program.
4) Changing economic conditions or ridership or cargo patterns may have affected the results of your program.

If any of these have occurred, the employment of a good statistician is mandatory.

Table 1
Tests for Detecting a Statistically Significant System Change

	Type of Evaluation	Conditions	Required Data	Statistical Evaluation Technique
a	**Determination if there has been some impact (non-quantified)**	The metric of interest has not been affected by anything other than the project. Except for the possible change in level caused by the project, the metric does not exhibit any trend, seasonality, other periodicity, or any other pattern or noticeable change.	A pre-implementation set of reference case metric data. The number of values in the reference case determines the significance level of the test One post-implementation metric value.	**Test A** Custom, distribution-free prediction limit test This test can be used to detect if a statistically significant change has occurred. It does not quantify the extent of that change.
b	**Determination if there has been some impact (non-quantified)**	The metric of interest has not been affected by anything other than the project. The metric may exhibit seasonality or other periodicity if these are of the same pattern and magnitude after implementation as before implementation.. The only other change in the metric is a possible change in level caused by the project.	Paired before-implementation and after fully operational values, where the elements of each pair come from the same place in any periodic cycle and as many of the stages of the cycle are represented as possible.	**Test B** Distribution-free, paired comparison tests These tests become more sensitive as the number of data pairs is increased. They indicate if the project has had a statistically significant impact on the metric, but do not quantify the extent of that impact.
c	**Quantified estimate of average impact**	The metric of interest has not been affected by anything other than the project. The metric does not exhibit any trend, seasonality, other periodicity, or change, other than a possible change in level caused by the project.	Period-by-period (e.g., monthly) metric values: at least 30 before project implementation and 30 after project fully operational	**Test C** One-sided, large sample test for a significant difference in means (averages).

<u>Table 1</u> <u>(continued)</u>

	Type of Eval	Conditions	Required Data	Stat Eval Technique
d	**Quantified estimate of impact**	The metric of interest has not been affected by anything other than the project **and** **Either:** (1) The metric exhibits no seasonality, other periodicity, or change except for the impact of the project and a possible linear trend, **or** (2) there is another variable or metric, such as traffic level, that historically has been highly correlated (proportional) to the metric of interest and thus can be used as a "control" or "predictor" variable. The metric of interest exhibits the same periodicity and trend as the control variable, except for the impact of the project.	Numerous values (e.g., monthly) of the metric and the control variable from before the project implementation and after the project is fully operational.	**<u>Test D</u>** Multiple Regression Analysis For Case (1): Regression against time and an indicator variable. For Case (2): Regression against a highly correlated predictor variable[2] and an indicator variable.
e	**Quantified estimate of average impact**	There is another variable such as traffic level, that historically has been highly correlated (proportional) to the metric of interest and thus can be used as a "quasi-control" or "gauge" variable because it should be unaffected by the implementation of the project.. The control variable exhibits the same periodicity, trend, or other pattern as that of the control variable, except for the impact of the project.	The metric of interest has not been recorded on a regular basis and its values (pre-implementation and post-operational) must be obtained through a focused study. Only a limited amount of data can be obtained.	**<u>Test E</u>** Impact Assessment Diagram technique (for use only when a limited amount of data can be obtained and other techniques cannot be used). Requires a "quasi-control"[4] or "gauge" variable.

[4] If the project is not implemented NAS-wide, one possibility for a "control" or "predictor" is "before" and "after" metric values for areas not impacted. The "metric of interest" values should then be for only those areas impacted by the project.

Table 1 (concluded)

	Type of Eval	Conditions	Required Data	Stat Eval Technique
f	**Best, most sensitive, quantified estimate of impact**	The metric of interest may exhibit a trend, seasonality, or other pattern. (This test has the least restrictive conditions.)	Regularly recorded, sequential data on all variables that might affect the metric of interest. At least 60 values of each variable with 1/3 to ½ of the values from the period after the project begins implementation. General knowledge of how the project will affect the metric as it becomes operational	**Test F** Box-Jenkins-Tiao Intervention Analysis with possible multivariate transfer function components. This requires a **good** statistician who is familiar with the technique.
x	**Quantified estimate of average impact based on predicted reference case metric values**	(1) There is good reason to believe that had the project not been implemented the reference case metric values for that period would have been worse than could be extrapolated from past reference case metric values, **and** (2) The original benefit analysis provided estimates of future reference case metric values **and** (3) The metric of interest has not been affected by anything other than the project, the metric exhibits no seasonality, other periodicity, trend, or change except for the impact of the project	At least 30 post-implementation period-by-period metric values and a single, average (mean), per-period reference case metric value.	**Test X** One-sided, large sample test for a mean value. The test compares the average of post-implementation metric values with the estimated average of what the values would have been without the project's implementation.

APPENDIX A
BENEFIT ANALYSIS
"RULES of CONDUCT"

The following rules and principles should be satisfied by any properly executed benefit estimation project. The number of rules may appear excessive, but they really are just common sense, and so should be reasonably easy to satisfy. As a benefit analysis progresses, it would be prudent to periodically review these rules and principles to ensure that the analysis is on track and to reduce the potential for later grief.

General Requirements

Guiding Principles

- Safety must not be compromised.
- There must be a documentable cause and effect (temporal) relationship between the investment and the benefits.
- Economic Benefits must be achievable in monetary terms by specific entities.
- Benefits should not be double-counted.
- Check for disbenefits that might result from the investment. For example, a project that increases terminal capacity also may have the potential of increasing the likelihood of a collision, particularly if it involves some technical risk.
- The documentation for each IA should include a complete description of the benefit estimation methodologies, the computations, and the data used.
- Documentation, data bases, and models should be retained for future use. Electronic versions should be archived so they don't disappear with departing staff or contractors.
- Plans for a post-implementation assessment of the actual benefits should be included in the IA, and should be implemented after the project is operational.

Reference case

- The reference case in year x should be "what the system would be in year x if we did not make this change."

Metrics Guidance

- The Metrics should be useable and measurable during modeling, operational trials and in-service operations.

- The Metrics should be in units of measurement that are useable in business cases by either or both Service Providers and Airspace Users

- Each metric should be clearly and completely defined. Any assumptions implicit in the definition of the metric should be made explicit and the potential ramifications of the assumptions should be described.

- Wherever possible metrics should be those already accepted. Other metrics should include a full explanation of the reasoning for their choice.

- There may be a choice of metrics available to measure a benefit category. (For example, for Safety one might use fatalities per million departures or fatal accidents per million flight hours.) In such cases, one should choose the metric most appropriate for the operational environment and project being studied. The ramifications of using other metrics should also be presented.

- If a metric (e.g., a safety metric) incorporates an exposure unit (e.g., flight hours, departures) as part of its definition, the definition and source of the exposure values shall be provided, and the ramifications of the use of different exposure units and any vagaries in the exposure values should be described.

<u>Quantification Guidance</u>

- Methods of measurement should, whenever possible, be objective and incorporate statistical methodology.

- If subjective methods of measurement are used for the quantification of a metric, they should not be the only measurement of that metric, and the subjective method should be adequately described and justified.

- Whenever different methodologies are used to quantify a metric in different phases of a program (e.g., modeling and operations), the relationships among the methods and the ramifications of the differences should be described to enable formal comparison of the measurements obtained.

- The source(s) of the data used to obtain the metric values, any deficiencies in the data, and algorithms for computing metric values shall be documented.

- For frequently used metrics and when possible, an easily accessed, current file should be maintained of the data used to generate the metric values.

- For frequently used metrics and when possible, the algorithm(s) used to generate the metric values should be automated.

- Wherever possible, the metric quantification methodologies should be based on those already developed.

APPENDIX B
The BENEFITS UNIVERSE

Looking for Benefits in all the Right Places

(A Benefit Categorization Scheme)

Any particular "type" of benefit can be viewed as being located in a four-dimensional "universe" of category "dimensions":

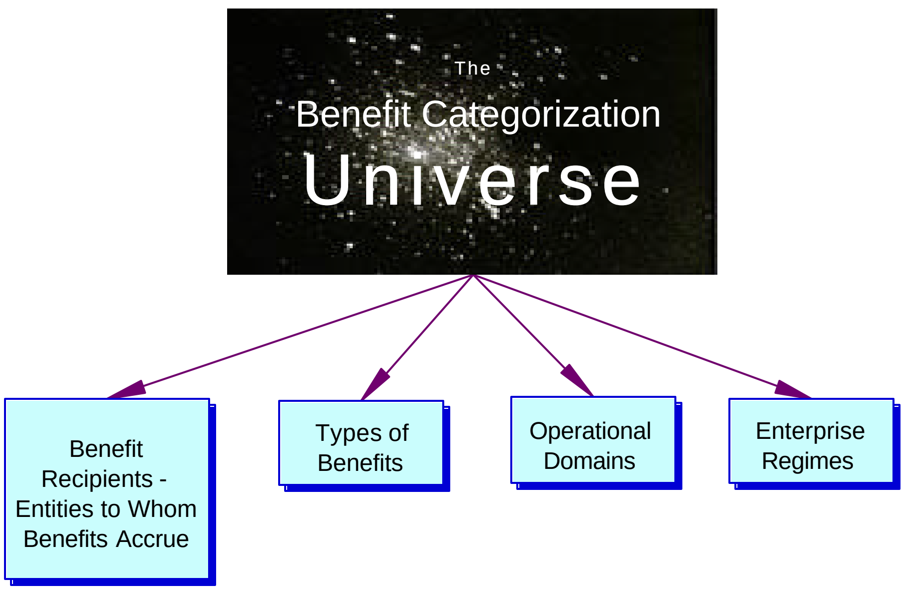

Figure B-1
The Benefits Universe

The above diagram, and the ones to follow, help clarify what needs to be done in a benefit analysis. Each of the four boxes in the second level of the above diagram is a "benefit category dimension." The use of these "dimensions" can best be described with an example.

Say we wish to estimate the safety benefits of a new system. The following diagram expands the *Types of Benefits* dimension to show the four possible locations along this dimension.

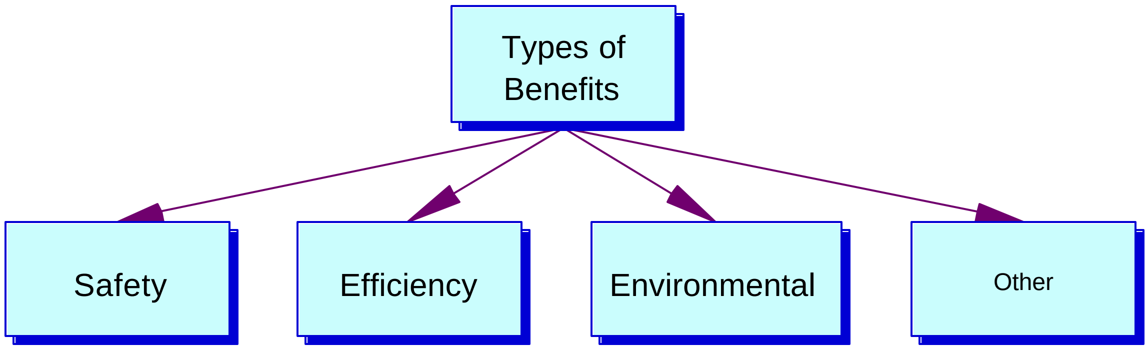

Figure B-2
Types of Benefits

Since we are only interested in safety benefits, it is obvious that our benefit category lies at the *Safety* location along the *Types of Benefits* dimension.

But what about the *Benefit Recipients* dimension? The next diagram depicts the locations of the entities along the *Benefit Recipients* dimension of our universe. Note that there are three main subdivisions of this dimension, that is, three classes of benefit recipients, namely the *Users of Services*, the *Providers of Services*, and *Society*. There are reasonably straightforward subdivisions of the first two of these; *Society* is more difficult to subdivide.

Where is our safety project located along the *Benefit Recipients* dimension? Won't it be located at several places along this dimension? That is, aren't there several entities that might benefit from improved safety? Depending upon the nature of our safety-enhancing project, passengers should benefit, air carriers may benefit, and general aviation may benefit: A project that would reduce collisions probably would benefit passengers, air carriers, and general aviation, whereas a project that enhances crash survivability might only benefit passengers.

In practice, we will have to separately estimate the benefits for passengers, air carriers, and general aviation. So, while the total benefits of our project may be distributed among more than one location along the *Benefit Recipients* dimension of our universe, our estimation of the benefits is done location-by-location.

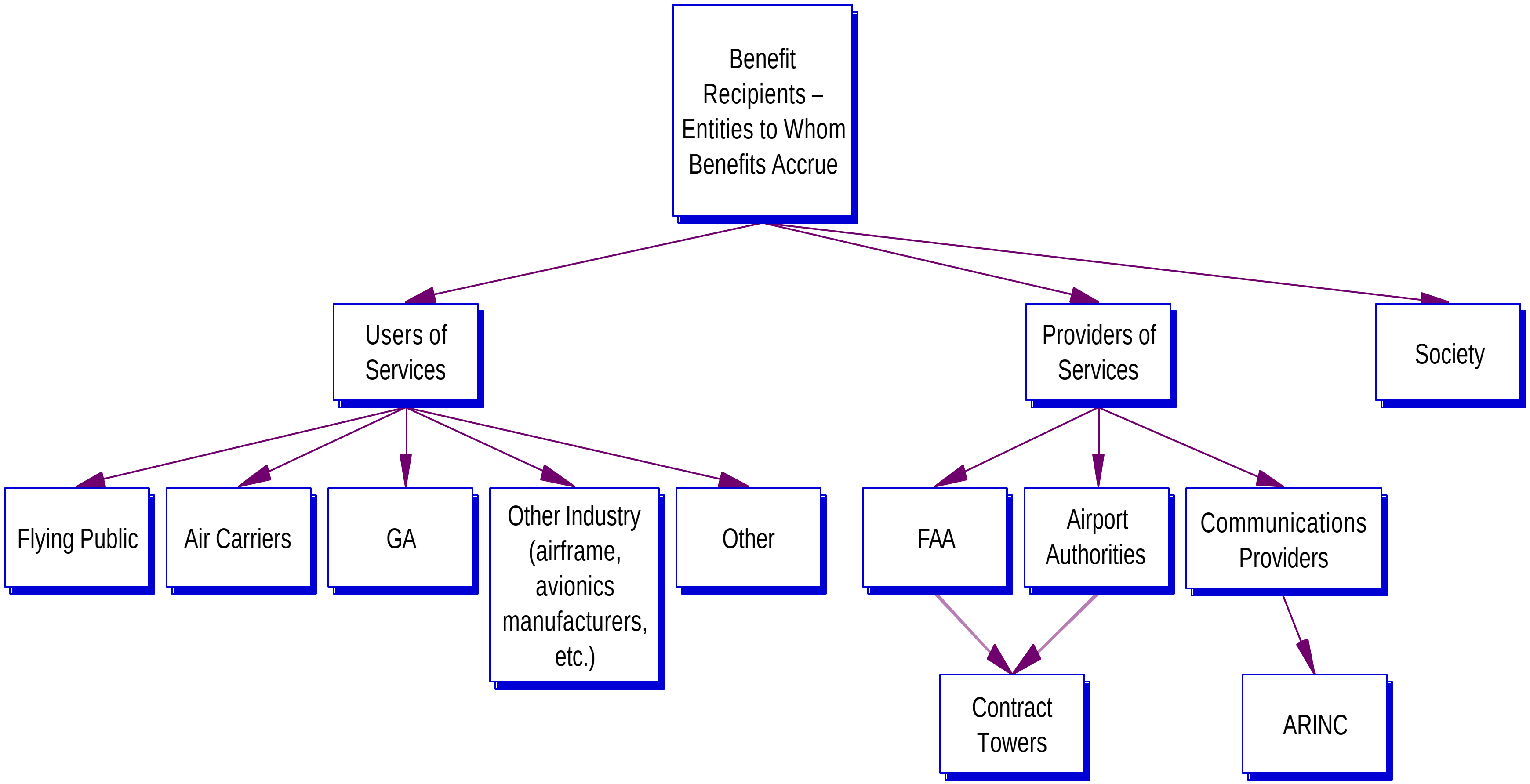

Figure B-3
Benefit Recipients

Knowing the type of benefits our project will have and the recipients of the benefits is not enough to begin the calculation of benefits. We must also include the physical environment(s) in which the benefits will occur. The next diagram illustrates the possibilities along the *Operational Domain* dimension.

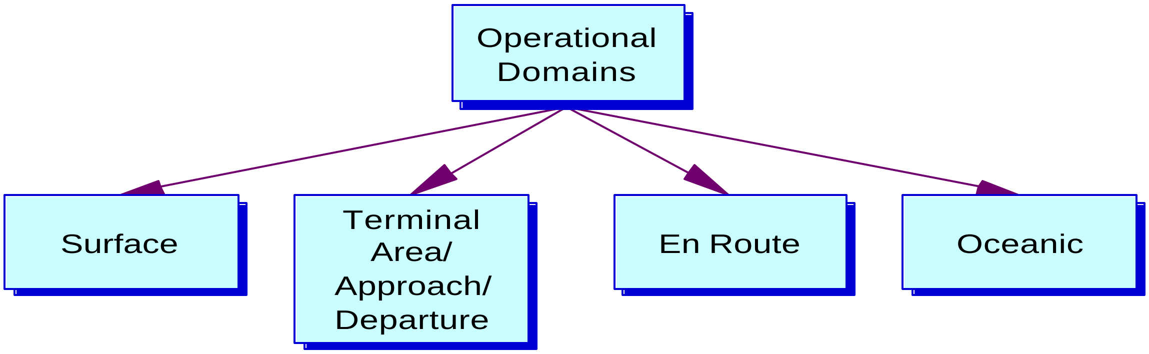

Figure B-4
Operational Domains

As before, because the nature of accident mechanisms may be different in each of these environments, we should separately calculate benefits at each location where they might occur.

A final dimension is related to the mechanisms of how the NAS operates and how our project will "physically" and operationally achieve its benefits. This dimension is called the *Enterprise Regimes*.

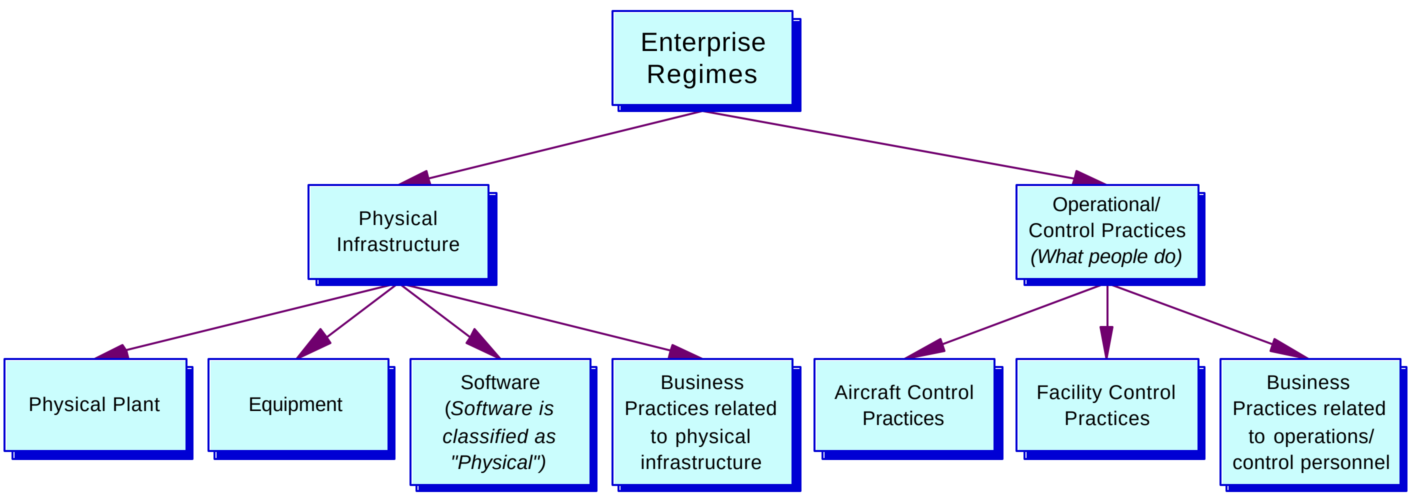

Figure B-5
Enterprise Regimes

APPENDIX C
METRICS - The LINK to BENEFITS

As described in the main text, an early step in the benefits evaluation process is an explicit description of the physical and operational effects of the project. These effects are usually measured in terms of changes in metrics. However, the metrics used are usually not ones that can be directly monetized.

For example, if our project includes ASDE-X, which identifies the positions of aircraft on runways and alerts the controller to potential runway collisions, a metric that is immediately impacted is the runway accident count or rate. However, we cannot immediately place a monetary value on a change in runway accidents. Instead, we must use a second set of metrics that are derived from the runway accidents metric. These metrics, which are monetizable, are passenger fatalities and injuries, and aircraft damage.

Thus, it seems that we need at least a two-tiered benefit metric structure. The first (and intermediate) tier(s) would be metrics measuring the improvements that would be directly realized from a project, while the final tier would be monetizable metrics derived from the preceding tiers' metrics. **The use of intermediate metrics helps clarify the determination of what the monetarily realizable benefits really are and aids in their quantification.**

Let's look at a second example that has been diagrammed in Figure C-1 to illustrate the process. Suppose a project is expected to both reduce delays at terminals and increase normal terminal capacity as the result of new technology. The first-tier metrics might then be chosen to be "Delay reduction" and "Capacity Improvement." Of course, there might be some overlap in counting these benefits, which is why the diagram below shows an overlap in "Capacity Improvement" and "Delay reduction."

Also shown, as a final tier of benefits, are <u>some</u> of the many benefits that might accrue from the successful implementation of this program. (Benefits not shown include, for example, "Reduction in Airline Revenue Loss.") We might be able to quantify and monetize some of these benefits but be unable to do so for others. Non-quantified benefits might still be worth discussing, however.

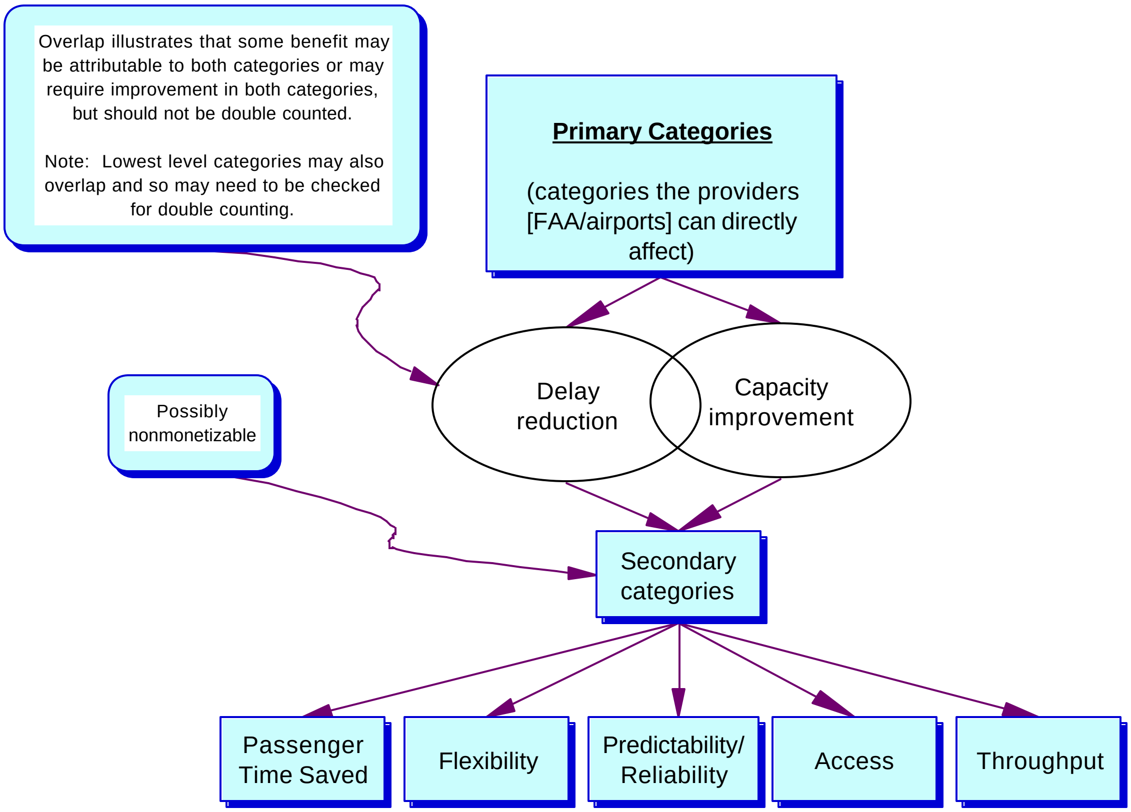

Figure C-1
Example Benefit Evaluation Process Categories

If our project will, say, improve the reliability of TRACON hardware and software, then one immediate impact will be a reduction in outages which, in turn, reduces delays. So, as shown below in Figure C-2, we can modify our diagram to show this intermediate, outage reduction metric. Depending on the nature of the project, there may be several intermediate metrics we will wish to include along the paths to the ultimate, monetizable metrics.

One should not think of the position of an item in the diagram as an indication of its level. Rather, one should think of an item as a stop along a road map. For example, in Figure C-2, "Outage Reduction" (metrics) is a useful stop on the path to reaching (obtaining) "Delay Reduction" (metrics).

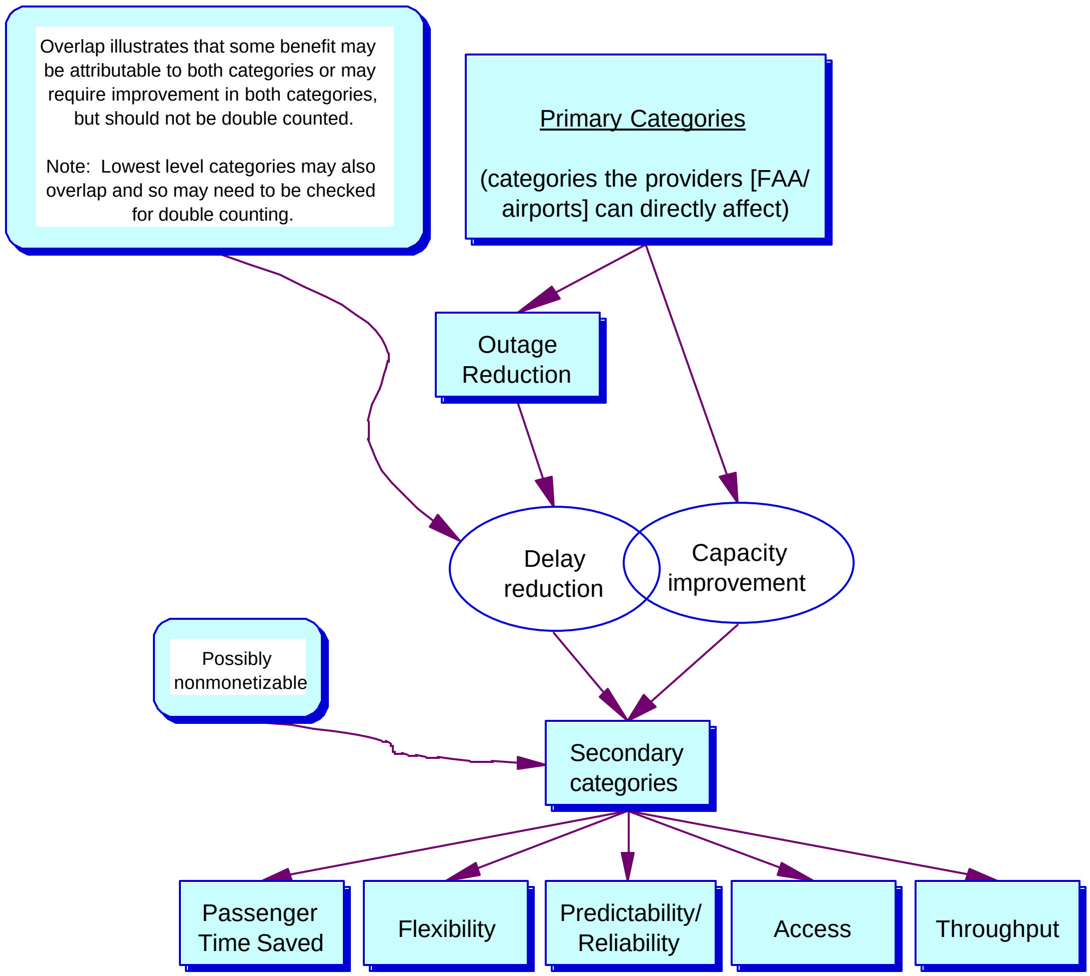

Figure C-2
Example Benefit Evaluation Process Categories ¾ Expanded

(this page intentionally blank)

APPENDIX D
DATA SOURCES

There are a multitude of data sources available for performing aviation-related analyses. In the following pages, we present brief descriptions of many of them, along with contact persons. Note however, that contacts can quickly become out of date. In particular, any listed SETA contacts are or shortly will be obsolete because of the transition to a new contractor. Corrections and revisions will be greatly appreciated.

DATABASE NAME	RESPONSIBLE OFFICE	DATABASE DESCRIPTION and CONTACT
ADA	APO-130	Aviation Data Analysis System - Includes Air Traffic Activity forecasts. Carlton Wine, 202-267-3350.
AFEIS		Air Facilities Executive Information System - Available to Division and Regional Managers. Contains outages and staffing information. Similar to EXIS. Rick Ford, AAF-60, 202-267-8970.
AFTECHNET		This web site contains daily reports on all scheduled and unscheduled outages that occurred in the NAS in excellent detail - http://aftechnet.faa.gov/ns.htm
ADOC		Airport Direct Operating Costs – Includes aircraft type and aircraft category costs by airborne hour and block hour costs. Data inputs are based on carrier submitted on Form-41.
ASAS		Aviation Safety Analysis System
ASQP	DOT	Airline Service Quality Performance - Developed to support a DOT report on airlines' on-time performance. Data elements include departure, arrival, and elapsed flight times as shown by (1) OAG, (2) carriers' reservations systems, and (3) carriers' actual performance. ASQP shows selected differences among the three sources, such as departure delay and elapsed time difference. However, it lacks the more detailed time and delay records of other databases. David Bennett, AAS-1, 202-267-3053. Gloria Laurie, DOT.
ASRS	ASY-200	Aviation Safety Reporting System - Contains operational errors, pilot deviations, and other air traffic problems voluntarily reported by pilots and controllers. ASRS data are used to identify deficiencies and discrepancies in the NAS so that these can be remedied by appropriate authorities, support policy formulation and planning for (and improvements to) the NAS, and strengthen the foundation of aviation human factors safety research. Tom Kossiaras, ASD-110, 202-358-5574.
ATADS	APO-110	Air Traffic Activity Data System – Provides operational count for Air Traffic Facilities. Nancy Trembly, APO-110, 202-267-9942.
ATOMS	ATM-300	Air Traffic Operations Management System – Provides regular count of air traffic operations and operations delays by minutes or more for all aircraft.
CBAS	ASD-420	Cost-Benefit Analysis System - Contains information on present and future costs and benefits of CIP projects to users and FAA. Brad Loomis, SETA, 202-651-2414.

DATABASE NAME	RESPONSIBLE OFFICE	DATABASE DESCRIPTION and CONTACT
CODAS	APO-130	Consolidated Operational Delay & Analysis System - A combined database of enhanced traffic management system (ETMS), airline service quality program (ASQP), and NOAA weather information. CODAS supports non-real-time analyses and projections of delays. Carlton Wine, APO-130, 202-267-3350.
COPS	ABC	Cost Performance System (COPS) - A data warehouse and decision support information system which allocates total FAA O&M appropriation costs to the field facilities, and associates these costs with workload and performance measures. Phillip Schaeffer, ABC-200, 202-267-9537 and ASD-430.
EDB		Engineering Data Base – End-state FAA system locations showing latitudes, longitudes, controlling ACF, antenna height, source/sink of functional interface, and specific subsystem connectivity. Terry Snyder, ARS-10, 202-366-9674 or Jim Novaco, SETA, 202-651-2271.
EIS	AAT	Air Traffic Executive Information System - Air Traffic version of EXIS. Larry Silvious, ATX-430, 202-267-7120.
ETMS	Volpe Center	Enhanced Traffic Management System - A database containing flights for which flight plans were filed and includes flight departure and arrival messages. It is available at the Volpe National Transportation Systems Center (Volpe Center) in Cambridge, MA. Tommie Tyson, AUA-500, 202-233-5052. Nancy Kalinowski – ATA-200.
EXIS	ABC-100	Executive Information System - Provides detailed concise demographic view of the FAA as compared with the national civilian labor force. Figures are broken down by line of business, as well as in terms hiring, promotions, and region. The Office of Business Information and Consultation updates information quarterly and at year's end. EXIS information is accessible to headquarters and regional management team members. Steve Hopkins, ABC-100, 202-267-7120.
F&E BSL	ASD-300	Facilities & Equipment Financial Reference case - Contains the financial reference case of F&E costs for current CIP projects. Dave Stuecheli, SETA, 202-651-2152.
FLAPS		FAA LINCS Architecture Pricing System – Provides the firm, fixed price cost of all Leased Interfacility NAS Communications System (LINCS) circuits and many other contract line item numbers (CLINs) for all ten years of the contract.
FMF &PFF	AOP-200	Facility Master File and Pre-Commission Facility File – Sub-element databases from the FSEP module of MMS, containing information on equipment and systems of FAA facilities from pre-construction through decommissioning. Ann Delaney, AOP-200, 202-267-3266 or Charlotte Powell, AOP-200, 202-267-3266.
FSEP	AOP-200	Facility, Service, and Equipment Profile - Database is described in FAA Order 6000.5C. It includes sub-elements, FMF and PCFF. Ann Delaney, AOP-200, 202-267-3266 or Charlotte Powell, AOP-200, 202-267-3266.
FSRDB	AND-140	Facility/Subsystem Requirement Database - Comprehensive listing of incoming CIP NAS subsystem component characteristics. The data elements collected include power, HVAC, environmental, dimensional and subsystem configuration data. Data on deployed CIP subsystems is migrated continually from the FSRDB to a separate but similar characteristics database as subsystems are installed fully. Dr. Sophia Ashley, AND-140, 202-358-5283.

DATABASE NAME	RESPONSIBLE OFFICE	DATABASE DESCRIPTION and CONTACT
LIS		LIS Engineering Database System - Maintains repair history for FAA Depot repaired items and maintains current information on modification records, performance data records, repair specification, manufacturer's information, and test equipment application. Ken Towery, Manager, NAILS Management Division, FAA Logistics Center, 405-954-4212 or Ellen Brinson, AND-340, 202-358-5040.
MMS		Maintenance Management System – All failure that have at least 1 minute duration, including NAPRS reports that have reliability and availability facility information by scheduled and unscheduled cause codes.
NAIMS	ASY-100	National Airspace Incident Monitoring System - Details of near mid air collisions, runway incursions, and causal factors. Bob Toenniessen, ASY-100, 493-4248 or Larry Randall, ASY-100, 493-4251.
NAPRS		National Airspace Performance Reporting System - Facility and services reports on scheduled and unscheduled outages, operational availability, operational delays and causes of delays. No longer considered a database. It is a set of requirements for what should be in Maintenance Management System (MMS). Frank DeMarco, AOP-200, 202-267-7359.
NASDAC		National Aviation Safety Data Analysis Center - Provides rapid automated access to a unique database that integrates commercial and government information, accident and incident data, aircraft-specific information, international safety recommendations, airport and navigational aids, and safety trend analyses. With a data storage capacity exceeding 300 billion bytes of information, the center houses one of the world's most extensive collections of aviation data. The center is staffed with analysts who are available to assist customers with NASDAC automation tools and data sources. FAA Headquarters, Room 1006, 800 Independence Ave. SW, Washington, DC, 202-483-4247.
NCDC	National Climatic Data Center	National Climatic Data Center database includes surface observation data, hourly weather updates of airports, and other useful aviation-related weather data.
NFDC	ATM-610	National Flight Data Center (OK City) – Contains "structural" information on the NAS, such as location of airports and navaids. Marie Killian, 202-267-5906.
NMNS	ASD-130	NAS Mission Need Statement Database – Source of information on description and status of every MNS throughout the FAA. Users of the database can view general information about the MNS (e.g. MNS Number, Title, Summary, and Status), as well as JRC and TSARC information (both past and future). Users may choose to print from a selection of existing reports. Gail Rollins, ASD-130, 202-358-4922.
NPIAS		National Plan of Integrated Airport Systems Database - Used by GAO to produce "Airport Development Needs Estimating Future Costs", Report No. GAO/CREDO-97-99 of April 8, 1997. Larry Kiernan, APP-400, 202-267-8784.
NTSB AAD	NTSB	NTSB Aviation Accident Database - Provides characteristics of all accidents, including the sequence of events, that occurred in the US airspace and summary narratives of each accident. Summary data available from Stan Smith. General telephone number, 202-314-6000; Public inquiries, 202-314-6551.

DATABASE NAME	RESPONSIBLE OFFICE	DATABASE DESCRIPTION and CONTACT
OAG	APO-130	Official Airline Guide - Official airport schedules of airline arrivals and departures. The OAG contains information on the flight's airline, flight number, arrival and departure cities, arrival and departure times, frequency of flight, connections, class of service, type of aircraft, number of stops and more. Gary Mihalik, 202-267-3347.
ODMS		Operational Data Management System
OPSNET	ATO-200	Operational Performance System Network – Used for air traffic delays and aircraft operations counts reporting. The planned evolution of the OPSNET is to include all radar terminal facilities and automated flight service stations (AFSS) and will include reporting requirements such as staffing and facility performance summaries. More information can be found in FAA Order 6040.15C (Titled: NAPRS). Larry Dixon, ATO-200, 703-925-3129.
PCFMF & PCPFF	AOP-200	PC versions of Facility Master File (FMF) and Pre-Commission Facility File (PFF). Ann Delaney, AOP-200, 202-267-3266 or Charlotte Powell, AOP-200, 202-267-5928.
PMAC	ASD-400	Performance Monitoring and Analysis Capability - A data analysis tool that provides accessibility to airline operations data in a PC environment. The PMAC system includes OAG, ASQP, CODAS, TAF, NCDC, and other data. Dan Citrenbaum, ASD-430, 202-358-5442. URL: http://www.faa.gov/opsresearch/pmac.htm
Reuters Aviation Database	Commercial	Reuters Aviation Database - Provides historical information from Airlines Form 41 fillings and the OAG. Allows for simple programming to create tables or database subsets of specific information from the Database. Includes operational, financial, personnel (e.g. number of flight crew, maintenance personnel, etc.) data.
RIMS	ARS	Requirements Information Management System - A comprehensive life cycle planning and data-tracking tool with four integrated modules: CIP Project Management, Budget Requirements Tools, Historical Cost, and Budget Planning. Rosanne Marion, ARR-200, 202-366-6934.
SDRS	ASY-100	Service Difficulty Reporting System - General aviation malfunction and defect reports and AC mechanical report. Bob Toenniessen, ASY-100, 202-493-4248 or James Hallock, VOLPE NTSC, 617-494-2199.
T-100 Airline Cost Data		Form 41 that includes carriers reporting costs by aircraft type – most of this information is applied by APO and reflected in FAA-APO-98-8, Economic Values for Evaluation of Federal Aviation Administration Investment and Regulatory Programs.
TAF	APO-110	Terminal Area Forecasts – 10-year forecasts of aviation activity at 873 airports in the U.S. by category of flight, i.e., air carrier, air taxi, general aviation. Dan Taylor, APO, 202-267-3302.
TIMS		Telecommunications Information Management System - Assists network planning, budget analysis, circuit engineering. Franklin Corpening, AOP-600, 202-267-9202.
TIS		Tower Information System - Provides graphical interface to "virtual database". Four paths to extract information: Airport, Equipment, Operations, and Tower. Information includes emplanements, tower details, future layout, current layout, runway list, runway details, equipment list, equipment details, equipment changes list, equipment changes details, equipment delivery, delays, operations, etc.

DATABASE NAME	RESPONSIBLE OFFICE	DATABASE DESCRIPTION and CONTACT
TTS+	AOP-100	Trouble Tracking System Plus reports failure/outage events from the NMCC for FAA facilities, a subset of the MMS – CSSI through AOP-100.
WIS	AFZ-200	Workload Information System - Provides maintenance staffing data for facilities. Barbara Froome, AFZ-200, 202-267-3203.

(this page intentionally blank)

APPENDIX E
MODELS & TOOLS

There are a multitude of available models for performing aviation-related analyses. In the following pages, we present brief descriptions of many of them, along with contact persons. Note however, that contacts can quickly become out of date. In particular, any listed SETA contacts are or shortly will be obsolete because of the transition to a new contractor. Corrections and revisions will be greatly appreciated.

MODEL NAME	DESCRIPTION and CONTACT
ABRM	The Analytic Blunder Risk Model is an analytic/probabilistic collision risk model programmed in Microsoft EXCEL. The model estimates collision risk for a given single-event scenario consisting of two aircraft under air traffic control: a blunderer (an aircraft deviating from a safe trajectory to one that crosses the path of another aircraft) and an evader (the threatened aircraft). Kenneth Geisinger, ATX-400, 202-267-8036.
ACEIT	Automated Cost Estimating Integrated Tool was developed for use in the Department of Defense. It uses a spreadsheet style structure to develop cost breakdown structures and contains an automated cost database of cost estimating relationships developed from industry data. Tecolote Research, Inc., URL: http://www.aceit.com/
ACIM	Air Carrier Investment Model - Generates estimates of the future demand for air travel from supply and demand factors based on projections of future economic conditions and operating characteristics of air carriers. Pete Kostiuk, Logistics Management Institute (LMI), 703-917-7427, pkostiuk@LMI.org
AEM	The Airspace Encounter Model was developed to estimate blind flying conflicts, collisions, and other encounters related to aircraft relative positions and velocities in NAS airspace. For example, AEM can be used to model aircraft conflict patterns under new concepts of operation. AEM can use the output of AOM to determine all potential conflicts among aircraft pairs occurring in a prescribed volume of airspace. AEM records the precise geometries of these conflicts, which can then be used in analyses of collision risk. Dr. Antonio A. Trani, VPI: 540/231-4418; FAX: 540/231-7532; vuela@vt.edu ; Stephen Cohen, ASD-430, 202-358-5230
AFCE	Airway Facilities Cost Estimating Model, a derivative of the Department of Defense's Cost Analysis Strategy Assessment (CASA) model, is specially tailored for use in estimating costs of FAA systems.
AIRNET	FAA Airport Network Policy Simulation Model is a queuing model that simulates a day's traffic through the US airport network. It allows users to see the impacts on airports, airlines, and passengers (in terms of time, dollars, noise levels) of airport capacity limitations and improvements, airport noise alleviation and access restrictions, and system performance under projected traffic growth. AIRNET addresses macro trends and interactions and calculates numbers for comparison to aid aviation-related policy planning and economic analyses. Carlton Wine, APO-130, 202-267-3350
Analytica	Develops complex influence diagram-based decision models and simulations. Mike Kaufman, SETA, 202-651-2293
AND	Approximate Network Delays is a quasi-analytical model of airport capacity and delay. Professor Amedeo Odoni, MIT, 617-253-7439, odoni@mit.edu or Dr. Andrew Haines, MITRE (CAASD), 703-883-6714, haines@mitre.org

MODEL NAME	DESCRIPTION and CONTACT
AOM	The Virginia Polytechnic Institute (VPI) Airspace Occupancy Model estimates three-dimensional airspace occupancies and provides input to the Airspace Encounter Model (AEM) AOM requires a series of aircraft flight plans and sector geometries as inputs. The model processes the information to determine the occupancy of each sector by different flights over time. The model stores the adjacency information of sectors, and identifies the sectors crossed by a flight plan. Dr. Antonio A. Trani, VPI: 540/231-4418; FAX: 540/231-7532; vuela@vt.edu; Stephen Cohen, ASD-430, 202-358-5230
ARC2000	Automatic Radar Control for the years beyond 2000 assesses the feasibility of automated ground-based separation assurance at a target date beyond 2015. ARC2000 demonstrates that automated air traffic control can maintain a conflict-free portion of the airspace for unlimited periods of time, and under high traffic densities. Xavier Fron, Eurocontrol, 011 33 1 69 88 75 30, fron.xavier@eurocontrol.fr or Jean-Pierre Nicolaon, Eurocontrol, 011 33 1 69 88 76 71, nic@eurocontrol.fr or Frederique Ayache, Eurocontrol, aya@eurocontrol.fr
ASAC	Aviation Systems Analysis Capability is developed for NASA to support Advanced Subsonic Technology (AST) Program. ASAC consists of several models: Air Carrier Investment, Airport Capacity, Airport Delay, Noise Impact, and Cost Models. Models and data repositories reside on the web and are accessible by FAA, NASA, and related industries. Peter Kostiuk, LMI, 703-917-7427
ASAT	The Airspace Simulation and Analysis for TERPS (ASAT) System is a multifaceted computer tool for aviation related simulations and evaluations. ASAT simulates various operational scenarios in realistic environments consisting of single or multiple aircraft, pilots and air traffic controllers. ASAT consists of high fidelity models and empirical data representing each component of real life scenarios, including aircraft, geographical, environmental, navigation systems, ATC systems and human factor models. ASAT uses these models to generate realistic aircraft positions in time and space and produces statistical data for risk analysis studies and visual representations. Alan B. Jones, AFS-420, 405-954-5844
ASCENT	ATFM System Concept Evaluator for New Technologies evaluates the system-wide impact of new procedures, technologies, and improved infrastructure under existing or anticipated future approaches to ATFM. Dr. Milton Adams, Draper Laboratory, adamsm@draper.com
ASIM	Airspace Simulation - Conflict resolution, workload measurement and airspace management. (British Civil Aviation Authority/National Air Traffic Services (CAA/NATS))
BDT	Banc De Test is a simulation tool that generates aircraft trajectories to test automated conflict resolution algorithms. Jean-Marc Alliot, Centre d'etudes de la Navigation Aerienne (CENA), 011-3362-17-4054, alliot@pc-allt.eis.enac.dgac.fr
CASA	Cost Analysis Strategy Assessment covers the life-cycle costs of the system, from initial research costs to those associated with yearly maintenance as well as spares, training costs, and other expenses once the system is delivered. Among the analyses it performs are production rate, quantity variation, warranty costs, operational availability, and several other related functions. CASA works by taking the data entered, calculating the projected costs and determining the probabilities of meeting, exceeding or falling short of any Life-Cycle Cost (LCC) target value. CASA offers a variety of strategy options and allows you to alter original parameters to observe the effects of such changes on strategy options. Ed Nedimala, ASD-410, 202-358-5220.
CheckPoint	Ed Begley, 617-273-0140
COCOMO	Constructive Costing Model estimates software development costs. The COCOMO Project Homepage is at URL: http://sunset.usc.edu/research/COCOMOII/index.html
COMNET	Network Simulation Model is a COMNET network-planning tool that includes COMNET Baseliner, COMNET lll, and COMNET Predictor. CACI Products Company; 3333 North Torry Pines Court; La Jolla, CA 92037; Phone 619-824-5200; E-mail: comnet@caciasl.com; URL http://www.caciasl.com/comnet.html

MODEL NAME	DESCRIPTION and CONTACT
COPS	Cost and Performance System is a prototype data warehouse and decision support information system that can allocate operations and maintenance (O&M) costs to field facilities. COPS can tie these costs to workload and performance data when measuring current costs of FAA facilities.
DELAYS	A dynamic queuing model that analyzes airport delays based on fleet mix, runway configuration, and demand.
DORATASK	UK CAA's Directorate of Operational Research and Analysis (DORA) - Sectorwise controller workload modeling. A fast-time simulation for evaluating sector capacity based on controller workload limits by systemically summing up the time the controller might spend on observable and non-observable tasks for each category of traffic in a sector. (CAA/NATS) Arnab Majumdar, Eurocontrol, arnab.majumdar@eurocontrol.fr
DPAT	Detailed Policy Assessment Tool. A national simulation model that predicts delays and measures performance for selected days as a function of parameters such as airspace and airport capacity. Provides delay metrics such as Departure Delay, Airborne Delay and Arrival Delay. Simulates 40,000 to 60,000 scheduled and unscheduled flights per day. MITRE/CAASD; Dan Citrenbaum, ASD-430, 202-358-5442
DPL	Software tool and programming language used to develop influence diagrams and decision trees for decision analysis. Mike Kaufman, SETA, 202-651-2293
ECOM	European Space Agency Cost Model is a software tool used for collecting, retrieving, and processing cost data from past ESA programs and projects. URL: http://www.estec.esa.nl/eawww/ecom/ecom.htm
EXPERT CHOICE	An analytic hierarchy process for multiple criteria decisions. Mike Kaufman, SETA, 202-651-2293
FAA Airfield Capacity Model	An analytical computer model which calculates the (maximum throughput) capacity of a runway system given continuous demand. William J. Swedish, CAASD, 703-883-6323
FLOWSIM	Daily Flow Simulation simulates the day's scheduled air traffic. Using traffic demand and airport capacity factors, FLOWSIM estimates how proposed traffic flow management strategies would affect the NAS. The model tests various planning options and displays the results graphically. The output includes a complete set of alternatives to help the traffic flow specialist resolve potential delay problems. John Bobbick, ATAC, 408-736-2822
FSM	Flight Schedule Monitor allows the traffic management specialist to examine (in real time) which airplanes are being moved in a Ground Delay Program. It also enables air traffic managers to visualize the airlines' flight cancellations and substitutions. Metron, 703-787-8700
GRADE	Graphical Airspace Design Environment is a state of the art, 4-D computer tool for displaying, analyzing, designing, and evaluating air traffic operations. Grade is a tool for airspace redesign, flight path and profile analysis, traffic flow, and sector loading analysis, obstruction analysis, environmental impact assessment, incident/accident investigation, and operational performance assessment.
HARS	High Altitude Route System is an automated traffic-planning tool that determines optimal flight routes based on aircraft performance, changing weather conditions, traffic demand, and resource limitations. HARS produces alternate route strategies for severe weather areas, special use airspace, or congested sectors. HARS is being used to aid ATCSCC planners in finding optimal re-routings around thunderstorms.
HERMES	Heuristic Runway Movement Event Simulation is a high-level of detail simulations of airport operations. It can be used to evaluate parallel runway or tower controller workload. David Haydon, 011 44 171 832 5601 (CAA/NATS)
HIPS	Conflict resolution, workload measurement and airspace management. Colin Meckiff, Eurocontrol, 331-6988-7601

MODEL NAME	DESCRIPTION and CONTACT
ICAO Collision Risk Model	The International Civil Aeronautics Organization (ICAO) provides guidance for separation analysis (e.g., [R6.6]). It has adopted a collision risk model developed by the North Atlantic System Planning Group (NAT SPG) to evaluate the safety implications of varying separation standards in the North Atlantic Oceanic Track System (NAT OTS). FAA contact: Brian Colamosca, ACT-520, 609/485-6603
INM	Integrated Noise Model is a regulatory model for determining annual noise impacts of airport operations. ASD-400 has worked closely with AEE to link SIMMOD to the INM to provide a relatively seamless and efficient airport tool set. INM will soon is able to calculate changes of exposure and population impacts within specified areas. The model is run on the Computer-Aided Engineering Graphics System (CAEGS). John Guilding, AEE-120, 202-267-3654
IWM	Integrated Wind Shear Model. Bob Juliano, SETA, 202-651-2419; Steve Cohen, ASD-430, 202-358-5230
LMI Runway Capacity Model	Generalized analytical and stochastic model for computing the capacity of a runway system. Its fundamental building block is a model that computes the capacity of a single runway, when the runway is used for arrivals only or for departures only or for mixed operations (arrivals and departures). Dr. David A. Lee, LMI, 703-917-7557, dlee@mail2.lmi.org or Dr. Peter F. Kostiuk, LMI, 703-917-7427, pkostiuk@lmi.org
Loral COTS Cost Model	Loral COTS Cost Model estimates the costs of COTS integration.
MIDAS	Man-Machine Integration, Design, and Analysis System. Kevin Corker, NASA/AMES, 650-604-0055, kevin_corker@qmgate.arc.nasa.gov
NARIM	National Airspace Resource Investment Model analyzes future airspace concepts. It is used to support FAA's and NASA's research and investment decision-making process, perform alternative analysis, determine impact of new procedures and technologies, and determine design requirements of new technologies. Diana Liang, ASD-430, 202-358-5236.
NARSIM	NLR ATC Research Simulator is a real-time Air Traffic Control simulation with humans and real ATC systems in the loop. It simulates aircraft, weather, and automated air traffic control. Nationaal Luchten Ruimtevaartlaboratorium, National Aerospace Laboratory, Netherlands (NLR). Michiels R., et. al., NARSIM Homepage, NLR
NASPAC	National Airspace System Performance Analysis Capability is a discrete-event simulation model that measures system performance. It tracks aircraft competing for air traffic control resources as they progress through the NAS. It enables the FAA and the aviation industry to study the effects of proposed changes in design, structure, and configuration of the various airspace and air components of the NAS. Dan Citrenbaum, ASD-430, 202-358-5442, Daniel.Citrenbaum@faa.gov
NASSIM	The NAS Simulation Model is a prototype engineering model used to support the NAS systems architecture definition process. It evaluates how the integrated components of the NAS impact each other, analyzes the embedded performance of proposed system enhancements, investigates alternate system designs or operational concepts, and evaluates impacts both from a system-level perspective and in high detail where required. Diana Liang, ASD-430, 202-358-5236
NIRS	Noise Impact Routing System provides optimization technology and methods in the TRACON and en route environments creating and evaluating alternatives for noise-minimum arrival and departure routes and procedures.
NOISIM	NOISIM is a real-time aircraft simulator with the ability to model and display the community noise impact of a specific trajectory that is flown. The model implicitly includes any aircraft-specific constraints and also includes the effect of wind or other atmospheric conditions on aircraft performance and noise propagation. John-Paul Clarke, MIT, 617-253-7748, johnpaul@mit.edu or Professor R. John Hansman, MIT, 617-253-2271, rjhans@mit.edu

MODEL NAME	DESCRIPTION and CONTACT
OPTIFLOW	Optimized Flow Planning is a decision support system for air traffic managers. It applies mathematical optimization techniques to generate air traffic management initiatives such as ground delay programs.
PBFM	Passenger and Baggage Flow Model is a discrete-event computer simulation model of the movement of passengers and baggage through an airport terminal. FAA William J. Hughes Technical Center.
PDARS	Performance Data Analysis and Reporting System is a means of capturing, storing, and analyzing SAR and ARTS radar track data.
PMAC	The Performance Monitoring Analysis Capability is a data analysis tool that provides accessibility to airline operations data in a PC environment. It supports several processes such as benefits analyses, mission needs analyses, performance metrics, model validation, etc. by providing analysts with a capability for better understanding National Airspace System (NAS) operations. Dan Citrenbaum, ASD-430, 202-358-5442, Dan.Citrenbaum@faa.gov.
PrecisionTree	Spreadsheet based decision tree development and analysis. Mike Kaufman, SETA, 202-651-2293.
PRICE-H	Parametric Review of Information for Costing and Evaluation, Hardware Model. It is used for deriving cost estimates of electromechanical hardware assemblies and systems. Earl Gillam, AUATAC, 202-314-1306
PRICE-HL	Parametric Review of Information for Costing and Evaluation, Hardware Life-cycle Model. It is used for deriving life cycle cost estimates of electromechanical hardware assemblies and systems. Earl Gillam, AUATAC, 202-314-1306. URL: http://www.pricesystems.com/
PRICE-M	Parametric Review of Information for Costing and Evaluation, Microcircuit and Electronic Module Model. It is used for deriving cost estimates of microcircuits. Earl Gillam, AUATAC, 202-314-1306. URL: http://www.pricesystems.com/
PRICE-S	Parametric Review of Information for Costing and Evaluation, Software Model Suite. It is used for deriving life cycle cost estimates software systems. Earl Gillam, AUATAC, 202-314-1306. URL: http://www.pricesystems.com/
PRICE-SL	Parametric Review of Information for Costing and Evaluation, Software Lifecycle Model. Earl Gillam, AUATAC, 202-314-1306. URL: http://www.pricesystems.com/
PUMA	Human factors; man-machine integration; workload model. Paul Day, Roke-Manor Research, paul.day@roke.co.uk
RAMS	Reorganized ATC Mathematical Simulator measures the workloads associated with ATC systems and organizations. It also offers users the possibility of carrying out planning, organizational, high-level, or in-depth studies of ATC concepts. Using multi-parameter conflict detection algorithms and an integrated rule-based conflict resolution system, RAMS offers the possibility of studying a wide range of ATC functions, from airspace management or route planning, to in-depth investigations of localized interest areas such as controller workload. (Eurocontrol, CACI). Diana Liang, ASD-430, 202-358-5236
RASRAM	Reduced Aircraft Separation Risk Assessment Model is a computer model used to assess the risk associated with aircraft operations. It measures the risk caused by reducing lateral or longitudinal separation and any subsequent reduction by introducing newer surveillance or navigation technology.
RATSG	Robust Air Traffic Situation Generator allows user to design 4D flight plans (position and time) for a number of pseudo aircraft for use in simulation studies. Professor John Hansman, MIT, 617-253-2271, rjhans@mit.edu.
REVIC	Revised Enhanced Version of Intermediate COCOMO. Air Force Cost Analysis Agency (AFCAA), 805-496-2505

MODEL NAME	DESCRIPTION and CONTACT
@ RISK	Risk Analysis and Simulation Add-In for Microsoft Excel or Lotus 1-2-3. A spreadsheet add-in tool used to conduct risk assessments using Monte Carlo or Latin Hypercube sampling techniques to simulate user-defined probability distributions of cost and benefits. Bob Juliano, SETA, 202-651-2419
Runway Capacity Model	Quasi-analytical models of airport capacity and delay. David A. Lee, LMI, 703-917-7557, dlee@mail2.lmi.org or Peter F. Kostiuk, LMI, 703-917-7427, pkostiuk@lmi.org
SASET	Software Architecture Sizing and Estimating Tool is similar to COCOMO that estimates the impact that software development will have on the schedule and cost of a program. Given certain information about the software code (such as the number of lines, whether it is new, modified, or reused, the complexity and the language), SASET will estimate how long it will take the project to go from design to end of Operational Test & Evaluation (OT&E) and the cost of the software development. Air Force Cost Analysis Agency (AFCAA), 805-496-2505
SATORI	Systemic Air Traffic Operations Research Initiative is an animation, simulation, and analysis tool used to recreate air traffic control operational incidents, review traffic management issues, develop facility-specific training programs, and present briefings on operational incidents. Mark Rogers, ASD-130, 202-358-5372
SDAT	Sector Design Analysis Tool provides 3-D design capabilities for sectors and traffic routes, calculates conflict potentials from air traffic samples to identify problem areas, and evaluates controller based on current and proposed sector design. Ken Geisinger, ATX-430, 202-267-8036
SEER-DFM	System Evaluation and Estimation of Resources - Design for Manufacturability. Earl Gillam, AUATAC, 202-314-1306 or GA SEER Technologies; Division of Galorth Associates, Inc.; 100 N. Sepulveda Blvd. - Suite 1801; El Segundo CA 90245; Phone 310-670-3404; E-mail: info@gaseer.com; URL http://www.gaseer.com/
SEER-H	System Evaluation and Estimation of Resources - Hardware. Earl Gillam, AUATAC, 202-314-1306 or GA SEER Technologies; Division of Galorth Associates, Inc.; 100 N. Sepulveda Blvd. - Suite 1801; El Segundo CA 90245; Phone 310-670-3404; E-mail: info@gaseer.com; URL http://www.gaseer.com/
SEER-HLC	System Evaluation and Estimation of Resources - Hardware Lifecycle. Earl Gillam, AUATAC, 202-314-1306 or GA SEER Technologies; Division of Galorth Associates, Inc.; 100 N. Sepulveda Blvd. - Suite 1801; El Segundo CA 90245; Phone 310-670-3404; E-mail: info@gaseer.com; URL http://www.gaseer.com/
SEER-IC	System Evaluation and Estimation of Resources - Integrated Circuit. Earl Gillam, AUATAC, 202-314-1306 or GA SEER Technologies; Division of Galorth Associates, Inc.; 100 N. Sepulveda Blvd. - Suite 1801; El Segundo CA 90245; Phone 310-670-3404; E-mail: info@gaseer.com; URL http://www.gaseer.com/
SEER-SEM	System Evaluation and Estimation of Resources - Software Evaluation Model. The SEER-SEM methodology is a sophisticated sizing and software estimating tool based on an extensive historic knowledge base, with over 800,000 million lines of code of completed software projects and a rich array of management trade-off capabilities. Earl Gillam, AUATAC, 202-314-1306 or GA SEER Technologies; Division of Galorth Associates, Inc.; 100 N. Sepulveda Blvd. – Suite 1801; El Segundo CA 90245; Phone 310-670-3404; E-mail: info@gaseer.com; URL http://www.gaseer.com/
SEER-SSM	System Evaluation and Estimation of Resources - Software Sizing Model. Earl Gillam, AUATAC, 202-314-1306 or GA SEER Technologies; Division of Galorth Associates, Inc.; 100 N. Sepulveda Blvd. - Suite 1801; El Segundo CA 90245; Phone 310-670-3404; E-mail: info@gaseer.com; URL http://www.gaseer.com/

MODEL NAME	DESCRIPTION and CONTACT
SIMMOD	FAA's Airport and Airspace Simulation Model evaluates airspace routing, airport expansion, hub-and-spoke operations, traffic demand and fleet mix, gate-taxiway-runway management, air traffic control procedures, and noise abatement procedures. Tony Vanchieri, ASD-430, 202-358-5198
SLIM	Software Life-Cycle Management is a sophisticated sizing and software-estimating tool based on an extensive historic database with over 4,400 completed software projects, with a rich array of management trade-off capabilities. Earl Gillam, AUATAC, 202-314-1306. URL: http://www.qsm.com/
SMARTFLO	Generates Traffic Flow Management (TFM) strategies for the ATCSCC by capturing the actual TFM Specialist's responses to daily flow situations and "learning" how experienced personnel handle various scenarios. SMARTFLO matches current conditions to similar past "experiences" and recommends intelligent strategies for managing traffic flow.
SODM	The System Outage Disruption Model (SODM) provides an easy way to estimate the effect on NAS system delay resulting from changes in the reliability and repair time of major FAA air traffic control (ATC) systems. The user provides the new reliability and repair time values and the future year being studied, and the model generates a probability distribution of total delay for that year relative to the 1997 baseline year. Steve Cohen, ASD-430. 202-358-5230
SPAS	Safety Performance Analysis System is a surveillance-planning tool for FAA safety inspectors and analysis. Barbara Wright, AFS-330, 202-267-7502
SPM	Spares Planning Model estimates fill rates at FAA inventory locations and the quantity of spares needed to achieve target fill rates. Thomas Pope, AFR-101, 202-493-0670
SoftCost-00	Resource Calculations, Inc., 303 267-0379
TAAM	Total Airspace and Airport Modeler is a high level of detail simulations of airport and airspace operations. The Total Airspace and Airport Modeller enables the evaluation of safety (conflicts and other separation infringements), capacity (number of movements, etc.), and economic effects (fuel flow and direct operating costs) of an Air Traffic Management (ATM) concept or airport design. TAAM uses a suite of analytical, model-based software modules and an advanced ATC simulation engine with powerful graphics. TAAM can randomly modify the traffic used in a simulation in order to test the scenarios for different traffic situations. Sasha Klein, B. Preston Group, 703-934-6190
Tactical TFM Testbed	Intermediate-level of detail simulations of airport and/or airspace operations. (Draper)
TAVT	Terminal Airspace Visualization Tool used for constructing, modifying, and displaying the complex terminal airspace in three dimensions. Designed specifically for the air traffic control application. (MITRE CAASD)
The Airport Machine	Tool for simulating in detail all aspects of airfield operations (including runways, taxiways, and apron areas). Its principal measures of performance (and outputs) are flows and throughput capacity on the airfield per unit of time, and delays experienced at the various airfield facilities. Ingrid Gerdes, (49)531 295 2279, ingrid.gerdes@dlr.de or Franz.Knabe, (49)531 295 2496, fllg@brzsp7.bs.dlr.de
TMAC	Traffic Flow Management Modeling and Analysis Capability. Intermediate-level of detail simulations of airport and/or airspace operations. John Pyburn, MITRE, 703-883-5546, jpyburn@mitre.org.

MODEL NAME	DESCRIPTION and CONTACT
TOPAZ	The Traffic Organization and Perturbation AnalyZer (TOPAZ) enables the evaluation of safety for a given (e.g., new) operational Air Traffic Management (ATM) concept during various flight phases. TOPAZ consists of a suite of analytical, model-based software modules, including a high-level Petri net-based simulation environment and mathematical packages to evaluate fatal ATM-related accidents. TOPAZ can incorporate probability estimates of rare deviations from normal operating conditions, which significantly distinguishes TOPAZ from commonly used, fast-time simulation environments, like the Total Airspace and Airport Modeller (TAAM). NLR, National Lucht-en Ruimtevaartlaboratorium, Amsterdam; Henk Blom, +31.205113544, blom@nlr.nl

APPENDIX F
STATISTICAL TESTS for POST-IMPLEMENTATION BENEFITS ASSESSMENT

A Word About Statistical Significance

We expect that the implementation of a project has caused an improvement in some part of the NAS[5] and thus a consequent improvement in those metric(s) that measure that part of NAS operations.

To determine this, we compare post-implementation values of the metric(s) with pre-implementation values of the metric(s). If we see improved values, we expect that these improved values are the result of the project. But how do we know that these improved values did not occur by chance? The way to make this determination is to perform a statistical significance test.

A statistical significance test is used to determine if we are justified in saying that the system has changed as a result of our project. This can perhaps best be explained in terms of an example.

> Suppose the goal of your project was to reduce a certain type of accident. Before the project was implemented about 20 of these accidents occurred each month. Some months there were a few more than 20 and some months there were a few less than 20. Once in a while there were many more than 20 and once in a while there were far fewer than 20. This month-to-month variation was usual and expected as a result of the monthly historical pattern of these accidents. (A statistician would call this variation *random variation* or *stochastic variation.*)
>
> Your project was implemented, and six months after it was fully operational, the GAO looked at the accident counts for the most recent three months. The monthly values were 12, 16, and 11. It seems that the project has had a positive effect, but can you "prove" to the GAO that the project really was beneficial. After all, in past years, there were a few months when the accident counts were this low. Maybe the project had no effect and, by "luck of the draw," these three months just "happened to have" low accident counts. The GAO wants ***proof***.

Statistics to the rescue!

> A statistical significance test can provide "proof" that there has been a beneficial change in the system. It does this by showing that the accident counts (12, 16, 11) are so small that it would be <u>very</u> unlikely that they would have occurred if there were no change in the system.

---

[5] National Airspace System

In the benefits assessment context, when we speak about a result being *statistically significant*, at say the 5% level, we essentially mean that **if** there were no change in the system, there would be then at most a 5% chance that we would obtain post-implementation metric values as good as (or better than) the values we obtained.

Note that this is not the same as saying that if our result is significant at the 5% level, there is one chance in 20 that the project did not improve the system. A 5% significance level says that there is **at most** one chance in 20 that the project did not improve the system. The actual chance may be much less, but we have no way of determining how much less.

By tradition, certain significance levels have become "standards." These are 0.1%, 0.25%, 0.5%, 1%, 2.5%, 5%, and 10%. The smaller the number, the more certain we are that an improvement actually occurred. If we can't obtain significance at even the 10% level, we usually are unwilling to claim with any certainty that there has been a positive impact. Note: Sometimes the word "confidence" is used (inappropriately) in place of the word "significance." In such cases, "**confidence level" = 100% - "significance level"** .

Also, it is important to distinguish between statistical significance and practical significance.

When a change is referred to as being statistically significant, all that is meant is that the data indicate that there has been *some* change ... that the change is greater than zero.

If one has little data, a test on that data might not yield a statistically significant result even though there is a substantial practical improvement. In this case, there just isn't enough data available to statistically detect the change or the statistical test used isn't sufficiently sensitive. Conversely, if one has a large amount of data and a sensitive test, one may obtain a statistically significant result based on a change that for practical purposes is insignificant.

If you wish to statistically determine whether there has been a practical change, you must first decide on how big the change has to be in order to be considered practical. That is, you must select a minimum value that must be achieved in order to say a change was practical. Then, a statistical significance test is used to determine if the post-change data are so different that it is very unlikely that the change was less than the minimum you selected. Statistical tests for a *practical* change are somewhat more complex than tests for *some* change. If you wish to test for a practical change, we suggest that you employ a statistician.

Statistical Software

There are many good statistical packages available. Even Excel® provides some statistical capabilities, although most statisticians have less than full trust in Excel® for statistical computations. The author personally prefers SPSS® (Statistical Package for the Social Sciences) for its accuracy, breadth, help screens, and remarkable ease of use. A less expensive statistical package that is also quite good, is reasonably priced, and is reasonably easy to use is NCSS® (Number Cruncher Statistical System). Minitab® is recommended by many statisticians, but it still does not include Fisher's Exact Test,[6] a serious omission in the opinion of this author, but apparently not in the opinion of its authors. SAS® (Statistical Analysis System) is generally considered the premiere statistical package. It's ease of use, however, leaves much to be desired (although it is, finally, improving) and its help screens often confound rather than elucidate.

The Statistical Tests

If possible, use the services of a good statistician to perform post-implementation assessments of a project's impact. A statistician will be able to craft tests more appropriate to a particular assessment than are presented in Table 1 and would be better able to determine if the conditions for the test are satisfied. If resources do not permit this, most of the tests in Table 1 can be performed by a good, non-statistician analyst. The following paragraphs describe how to conduct all but one of these tests. The tests are presented in "cookbook" form. Before using a test, **be sure all of the conditions listed for the test in Table 1 are satisfied**, and **be sure you are NOT using moving average data.**[7]

Also, **you are not allowed to repeatedly try a test with new post-implementation metric values until you get a statistically significant result.** By repeatedly testing with new data, you are no longer working with the same significance levels. However, it is permissible to try different tests with the same data, provided the conditions for the tests are met.

If you need to test the post-implementation effectiveness of a project before it is fully operational throughout the NAS, you can test by restricting your "universe" to where the project is fully operational and using metric values from only that "universe."

Finally, **plot your data.** The human eye-brain combination can often recognize patterns and anomalies that escape detection by formal analytic methods.

[6] You may have occasion to need Fisher's Exact Test.
[7] Moving average values are highly autocorrelated which makes them unsuitable for these tests.

Test A: Custom, distribution-free prediction limit test

This distribution-free test is extremely simple to use. It's derivation, which is not so simple, is based on combinatorics and the use of the hypergeometric distribution function. (For further details, see the test developer, Steve Cohen, ASD-430.) For ease of explanation, we will describe the test assuming the metric values are monthly. This is only for descriptive purposes, the metric values need not be monthly, but must be measurements from equal periods of time.

Step 1: Be sure all of the conditions listed for the test in Table 1 are satisfied and that your data are not moving average values.

Step 2: Choose a baseline period with monthly metric values representative of the system before the implementation of the project. List these metric values in chronological order. The more values in this base period, the greater the possible significance level of the test. (No fair cheating by picking particularly "bad" pre-implementation metric values!)

Step 3: The significance level of the test depends on the number of metric values in the representative baseline:

Number of baseline metric values	Significance level
at least 9	10% or better*
at least 19	5% or better
at least 39	2.5% or better
at least 99	1% or better

* "Better" means a smaller number: For example, 5% is better than 10%. That is, 5% is more significant than 10%.

Step 4: Choose a month after the project is fully operational and choose as a test value the metric value **T** for that month. (Once you have chosen a month, it is invalid to choose a "better" month to get "better" results.)

Step 5: The test proceeds iteratively, attempting to find increasingly better significance levels using increasing numbers of baseline metric values. At each iteration,

a) If smaller metric values indicate improved performance, compare **T** with **S**, the smallest metric value in the baseline period. If **T < S** , then the test has yielded statistically significant results at the level shown above for the number of baseline metric values used.

b) If larger metric values indicate improved performance, compare **T** with **L**, the largest metric value in the baseline period. If **T > L** , then the test has yielded statistically significant results at the level shown above for the number of baseline metric values used.

Step 6: The iteration begins here.

a) Select the first 9 metric values in your baseline period[8] and perform the appropriate comparison in **Step 5.** If the test is passed, you have determined, with a statistical significance level of 10%, that there has been an improvement. That is, you can state that, "a 10% significance level statistical test indicates that there has been an improvement in the metric." Proceed to Step 6(b).

If the test is not passed, you cannot state, at even the 10% significance level, that there has been an improvement. Stop.

b) Select the first 19 metric values in your baseline period (if you have that many) and perform the appropriate comparison in **Step 5.** If the test is passed, you have determined, with a statistical significance level of 5%, that there has been an improvement. That is, you can state that, "a 5% significance level statistical test indicates that there has been an improvement in the metric." Proceed to Step 6(c).

If this test is not passed, you do not have significance at the 5% level, but you do have significance at the 10% level or better. Stop.

c) Select the first 39 metric values in your baseline period (if you have that many) and perform the appropriate comparison in **Step 5.** If the test is passed, you have determined, with a statistical significance level of 2.5%, that there has been an improvement. That is, you can state that, "a 2.5% significance level statistical test indicates that there has been an improvement in the metric." Proceed to Step 6(d).

If this test is not passed, you do not have significance at the 2.5% level, but you do have significance at the 5% level or better. Stop.

d) Select the first 99 metric values in your baseline period (if you have that many) and perform the appropriate comparison in **Step 5.** If the test is passed, you have determined, with a statistical significance level of 1%, that there has been an improvement. That is, you can state that, "a 1% significance level statistical test indicates that there has been an improvement in the metric."

If this test is not passed, you do not have significance at the 1% level, but you do have significance at the 2.5% level or better.

---

[8] It is assumed that choosing the first 9 values is equivalent to randomly choosing 9 values from the entire baseline period because, by the rules in Table 1, there is no pattern in the metric value data.

Example: Suppose in Step 2 the base period you chose as a representative of the pre-implementation period included the following 22 metric values:

263, 316, 276, 414, 333, 257, 312, 289, 274, 308, 264,
317, 288, 249, 279, 302, 337, 324, 292, 241, 318, 299.

Suppose for the project under study smaller metric values indicate improvement. Also, suppose that after the project is fully operational we choose a month and the metric value for that month is **T = 251**.

The first 9 of the baseline values are 263, 316, 276, 414, 333, 257, 312, 289, 274. The smallest of these values is **S = 257**. The largest of these values is **L = 414**. Because "smaller values are better.", we compare **T** with **S**. **T < S**, so we can state that, "we have shown that a 10% significance level statistical test indicates that there has been an improvement in the metric."

Next, we choose the first 19 values. These are, 263, 316, 276, 414, 333, 257, 312, 289, 274, 308, 264, 317, 288, 249, 279, 302, 337, 324, 292. The smallest of these values is **S = 249**. In this case , **T > S**, so we have not achieved statistical significance at the 5% level.

Test B: Distribution-free, paired comparison tests

There are two tests offered here. The first, more sensitive test is based on an assumption regarding the data that is not required for the second test. Both tests begin with the same steps. (Note: These tests have been slightly modified to better match the properties of FAA metric data.) For ease of explanation, we will describe the tests assuming the data values are monthly. This is only for descriptive purposes, the data values need not be monthly. These tests can be found in some elementary statistics books and in virtually all nonparametric statistics books.

Step 1: Be sure all of the conditions listed for the test in Table 1 are satisfied and that your data are not moving average values.

Step 2: Randomly select a sample of monthly metric values from the time period before implementation of the project. If the metric values are seasonal or otherwise periodic, try to select values representing different parts of the seasonal (periodic) cycle.

From the time period after the project has been fully implemented, select a one-for-one corresponding sample of monthly metric values. For example, if your before-implementation sample includes February 1997 and February 1999 values, then your after-implementation sample should include two February values that are two years apart.

Step 3: Match in pairs and in chronological order the before-implementation metric values and the after-implementation values. For instance, one might have as matched pairs

(2/1997 value, 2/2003 value), (5/1997 value, 5/2003 value)
(11/1997 value, 11/2003 value), (2/1999 value, 2/2005 value) , etc.

(This style of matching should reduce the confounding effects of any seasonality in the data.)

Step 4: For each data pair $(\mathbf{b_i},\mathbf{a_i})$ ($\mathbf{b_i}$ a "before" value, $\mathbf{a_i}$ an "after" value) find the difference in the two values,

$$\mathbf{D_i = b_i - a_i} \quad (\Leftarrow \text{Order is important: BEFORE} - \text{AFTER}).$$

If $\mathbf{D_i = 0}$, remove the corresponding $(\mathbf{b_i},\mathbf{a_i})$ pair from the data.

Step 5: In this step you determine whether the more sensitive test can be used. (The procedure presented here is an informal, inspection technique rather than a formal statistical procedure.)

a) Arrange the differences $\mathbf{D_i}$ in ascending order.

Example: -70, -40, -10, 30, 50, 60.

b) Find the median **M** of these values. (**M** is the middle value if the number of differences is odd; **M** is the average of the two middle values if the number of differences is even.)

For the example, $\mathbf{M} = (-10+30)/2 = 10$.

c) Calculate the values $\mathbf{E_i = D_i - M}$. That is, subtract **M** from each $\mathbf{D_i}$, retaining the ascending order in the results.

Example: Using the values in part (a), $\mathbf{E_i}$ = -80, -50, -20, 20, 40, 50

d) Now, ask yourself, "Are these values distributed in a reasonably symmetric way about the number **0**? If you are not sure, match the smallest positive number with its negative counterpart, which is the negative number closest to the value **0**. Do the same with the rest of the values.

Example:		
	20	-20
	40	-50
	50	-80

In this example, the values are not reasonably symmetric as the magnitudes of most of the negative numbers are larger than their positive counterparts.

e) If there is reasonable symmetry, use test $\mathbf{B_1}$. If there is not reasonable symmetry, or you are uncertain, use test $\mathbf{B_2}$.

Test B_1: The one-sided Wilcoxon Matched-Pairs Signed Ranks Test.

Step B_1-1: Rank, from smallest to largest, the **magnitudes** (i.e., absolute values) of the values $\mathbf{D_i}$ that were calculated in Step 4**.**

Then, attach to each rank value the sign (**+** or **-**) of the corresponding original difference.

Ties: If two or more of the magnitudes are equal, give each the average of the ranks that otherwise would have been assigned to them.

Example: (We will use for illustration only the same numbers that were used in Step 5(a), above, although they really do not satisfy this test's conditions.) [Careful! Don't accidentally use the $\mathbf{E_i}$ values.)

Magnitude of D_i	Rank	Signed Rank
10	1	-1
30	2	+2
40	3	-3
50	4	+4
60	5	+5
70	6	-6

Step B_1-2: Compute $\mathbf{T_+}$ = the sum of the ranks with positive signs.
Compute $\mathbf{T_-}$ = the sum of the ranks with negative signs.

Example: Continuing with the example in Step B_1-1,
$\mathbf{T_+}$ = +2 +4 +5 = 11
$\mathbf{T_-}$ = 1 + 3 + 6 = 10 (← The sum of the ranks that are "-")

Step B_1-3: If smaller metric values indicate improvement, use $\mathbf{T_-}$ **.**
If larger metric values indicate improvement, use $\mathbf{T_+}$.

Using the appropriate **T** value, and the number **N** of data pairs (after eliminating those with a 0 difference), use Table F-2 to determine if the test indicates that there has been a statistically significant improvement in the metric.

Here's how to do it:

a) Find your value of **N** in the left column of the table. If **N** > 30, see the procedure below.

b) Opposite the value of **N** , look across the table for the column containing your value of **T** . If your value of **T** lies between two numbers in the table, pick the column containing the larger of these two numbers (the one to the left).

c) The significance level a is at the top of the column you picked.

d) If your value of **T** falls within the table, you can state that, "an a **%** significance level statistical test indicates that the project has had a beneficial impact."

e) If your value of **T** is smaller than any of the numbers in the row for **N** , your significance level is better than 0.5%, a very good result.

f) If your value of **T** is larger than any of the numbers in the row for **N** , your significance level is worse than 10%, so this test does not confirm that there has been any improvement in the system as a result of your project.

Examples:

i) Continuing with the example in Step B_1-2, if larger metric values indicate improvement, we use $\mathbf{T_+ = 11}$. $\mathbf{N = 6}$, so we use the first row in Table F-2. $\mathbf{T_+}$ is (much) larger than the value **4** in the table so this test certainly does not yield evidence that there has been any improvement in the system.

ii) Suppose in a second example, smaller metric values indicate improvement, and in this example **N = 14** and $\mathbf{T_- = 19}$. In Table F-2, opposite **14** in the **N** column we see that our value of $\mathbf{T_- = 19}$ falls between columns containing the numbers 21 and 16. So we choose the column to the left (containing 21). At the top of that column, we read the significance level of 2.5%. We therefore can say that, "a 2.5% significance level statistical test indicates that the project has had a beneficial impact."

When N > 30, use the following procedure.

Calculate the value

$$\mathbf{z = 1.2247 \ x \ \frac{4T\text{-}N(N+1)}{\sqrt{N(N+1)(2N+1)}}}$$

Compare this value of **z** with the values in Table F-1 to obtain the statistical significance level.

Test B_2: The One-Sided Sign Test.

In Step 4 (above Test B_1), you calculated for each data pair $(\mathbf{b_i},\mathbf{a_i})$, the difference in the two values, $\mathbf{D_i = b_i\text{-}a_i}$, and eliminated any pair for which $\mathbf{D_i = 0}$.

Step B_2-1:

If smaller metric values indicate improvement, count the <u>number</u> of <u>negative</u> $\mathbf{D_i}$ values. Call this count **C**.

If larger metric values indicate improvement, count the <u>number</u> of <u>positive</u> $\mathbf{D_i}$ values. Call this count **C**.

<u>Example</u>: If the $\mathbf{D_i}$ values are -7, -4, -1, 3, 5, 6, 8, 12 , then
C = 3 if smaller metric values indicate improvement, and
C = 5 if larger metric values indicate improvement

Step B_2-2: Let **N** be the total number of pairs with non-zero differences $\mathbf{D_i}$. In Table F-3, find the value in the table corresponding to **N** and **C**. If **N > 30**, see the procedure below.

The value in the table is the probability of obtaining the specified **C** value (or smaller) if, in fact, the project had no effect, or worse, had a deleterious effect.

To quote a standard significance level for the test, choose the smallest value in the following list that is at least as large as your table value

0.1%, 0.25%, 0.5%, 1%, 2.5%, 5%, 10%.

For instance, if the table value is 0.021, the standard significance level is 2.5%. (If desired, you can instead use the actual value in the table, stating that the significance level is 2.1%.)

Examples:

For **N=14** and **C=4**, the value in Table F-3 is 0.090 = 9.0%. The corresponding standard significance level is 10%.

For **N = 8** and **C = 3**, the value in Table F-3 is 0.363. So the probability of obtaining a value of **C = 3** (or smaller) when one has N **= 8** pairs of data if the project did not improve the system is p = 0.363 = 36.3%. This value is much larger than any of the standard significance levels. Thus, in this example these data do not furnish evidence that the project is beneficial.

When N > 30, use the following procedure.

Calculate the value

$$\mathbf{z} = \frac{\mathbf{2C - N + 1}}{\sqrt{\mathbf{N}}}\ .$$ [9]

Compare this **z** value with the values in Table F-1 to obtain the significance level.

Test C: One-sided, large sample test for a significant difference in means (averages).

This is the usual test for a difference in population means for the case where the two populations may have different variances. The test may be found in any elementary statistics textbook. For convenience we describe the test in terms of monthly metric values, but regularly recorded metric values for some other period (e.g., weekly, daily, etc.) can also be used.

Step 1: Be sure all of the conditions listed for the test in Table 1 are satisfied and that your data are not moving average values.

Step 2: Randomly select a sample of at least 30 monthly metric values from the time period before implementation of the project, and a sample of at

[9] Regarding the **+1** in the equation, Siegel, pg. 72 (see footnote on Table F-3) and others advise the use **+1** if **C < N/2** and **–1** if **C > N/2** . However, the author has checked numerous cases when **C > N/2,** and in every case **+1** gives a much more accurate answer.

least 30 metric values from the time period after the project is fully operational.

Let $\mathbf{n_b}$ denote the number of metric values in the pre-implementation (before) sample, and let $\mathbf{n_a}$ denote the number of metric values in the post-implementation (after) sample.

Step 3: Calculate the mean $\bar{\mathbf{X}}$ and variance $\mathbf{V_b}$ of the pre-implementation sample metric values $\mathbf{x_i}$:

$$\bar{X} = \frac{\sum_{i=1}^{n_b} x_i}{n_b}, \qquad V_b = \frac{\sum_{i=1}^{n_b} (x_i - \bar{X})^2}{n_b - 1}$$

Calculate the mean $\bar{\mathbf{Y}}$ and variance $\mathbf{V_a}$ of the post-implementation sample metric values $\mathbf{y_j}$:

$$\bar{Y} = \frac{\sum_{j=1}^{n_a} y_j}{n_a}, \qquad V_a = \frac{\sum_{j=1}^{n_a} (y_j - \bar{Y})^2}{n_a - 1}$$

Step 4: **a)** If smaller metric values indicate improvement and $\bar{\mathbf{Y}} < \bar{\mathbf{X}}$, continue with Step 5.

b) If larger metric values indicate improvement and $\bar{\mathbf{Y}} > \bar{\mathbf{X}}$, continue with Step 5.

c) If neither of the above is true, stop. Either the program has not been beneficial or some other factor has prevented improvement.

Step 5: Calculate the value

$$z = \frac{-\left|\bar{X} - \bar{Y}\right|}{\sqrt{\frac{V_b}{n_b} + \frac{V_a}{n_a}}}$$

Step 6: Compare this **z** value with the values in Table F-1 to obtain the significance level.

Step 7: If Step 6 results in a statistical significance level of 10% or better, you can state that, "a __% significance level test indicates that the project improved the system with an average monthly improvement of (the value) $\left|\bar{X} - \bar{Y}\right|$."

Test D: Multiple Regression analysis with an indicator variable

(or Regression analysis with a highly correlated predictor variable and an indicator variable).

This test requires you to know how to run a regression. Regression analysis is available in Excel®[10] and all general-purpose statistical software packages. You may need to use a control or predictor variable[11]. A discussion of regression may be found in all elementary statistics books.

Step 1: Be sure all of the conditions listed for the test in Table 1 are satisfied and that your data are not moving average values.

Step 2: By carefully inspecting a plot against time of the pre-implementation metric values, determine if the metric **M** exhibits seasonality, other periodicity, or some other non-trend pattern.

a) If **M** does not exhibit any pattern, other than possibly a trend, choose time as the control variable **C** in Step 3, below. Time should be expressed as sequential numbers (e.g., 1 for the first month, 2 for the second month, etc.).

b) If the metric values exhibit seasonality or any other non-trend pattern you have to find a control (predictor) variable **C** that "explains" all of the pattern in the metric **M** except for any impact due to the impact of the project.[12] If you can't find such a variable **C**, then don't use this test..

Step 3: Select sequential pairs $(\mathbf{C_i},\mathbf{M_i})$ of control[13] and metric values from the time period before the project was implemented.

Select sequential pairs $(\mathbf{C_i},\mathbf{M_i})$ of control and metric values from the time period during which the project was fully operational.

(There usually will be a break between the "before" and "after" period during which the project was in the process of being implemented.)

Step 4: In this step you determine if the control variable **C** you chose is a good choice. If (except for a level change due to implementation of the

---

[10] Note, however, that most statisticians have less than full trust in Excel® for statistical computations. Excel® is a registered trademark of the Microsoft Corporation.

[11] If the project is not implemented NAS-wide, one possibility for a "control" or "predictor" is "before" and "after" metric values for areas not impacted. The "metric of interest" values should then be for only those areas impacted by the project.

[12] Ibid.

[13] Ibid.

project) the metric does not exhibit **any** pattern, not even a trend, then you can skip this step.

Using only the pre-implementation data pairs, regress **M** against **C**. If **C** is a good control (predictor) variable for the metric **M**, then the regression should produce <u>all</u> of the following results. If it does not, do not use this test.

a) An $\mathbf{R^2}$ value of 0.90 or greater (= a correlation value of **R** = +0.95, or better). The regression analysis software you use should provide you with the value of $\mathbf{R^2}$ or of **R**. If it does not, find better software.

b) The estimated coefficient of **C** is **positive** and is statistically significant at the 10% level or better (or equivalently, the "confidence level" is $\geq$ 90%). If your software does not provide information about the statistical significance of the coefficient of **C**, junk it. It is worthless.

c) The residuals from the regression should reasonably follow a normal distribution. Hopefully, your regression software either provides a numerical test result of this or it provides a normal probability plot of the residuals so you can tell by inspection if the residuals are reasonably normally distributed. (Note: Many statistical packages (and Excel®) require you to specifically request (check a box) normality information before you run the regression.)

The way to use the normal probability plot is to determine, by inspection, if all but at most one or two of the points on the plot reasonably fall along a straight line. If the plot includes a straight line and 5% or 10% "bounding curves" on either side of the line, then no more than 5% or 10%, respectively, of the points can be outside the bounding curves.

Step 5: Create an *indicator variable* **L** as follows:

$\mathbf{L_i = 0}$, if the data point $\mathbf{(c_i,m_i)}$ is from the before-implementation period.
$\mathbf{L_i = 1}$, if the data point $\mathbf{(c_i,m_i)}$ is from the post-implementation period.

L represents the impact of the project, that is, the change in the level of the metric **M** after the project is fully operational.

Step 6: Using <u>all</u> of the data points $\mathbf{(C_i,L_i,M_i)}$, perform a multiple regression, regressing **M** against both **C** and **L**.

Step 7: Check the statistical significance level of the coefficient l of the indicator variable **L** in this multiple regression. (Your software should provide this value.) If this significance level is 10% or better (or equivalently, the "confidence level" is $\geq$ 90%), you can state that "the data indicate that the project resulted in a change of (size) l in the level of the metric **M**."

Test E: Impact Assessment Diagram Technique.

This is a special technique for use when metric data has not been regularly collected, it is necessary to do a focused study to obtain relevant data, and none of the usual, standard statistical techniques are applicable. The existence and use of a "quasi-control" or "gauge" variable is required. This technique is not in statistics books, so we will spend some time describing it. Also, it should be noted that this is not a particularly sensitive test and it should not be used if other tests are applicable. For further information see the test developer, Steve Cohen, FAA/ASD-430.

In order to determine if this technique is applicable, the ideas behind it need to be explained.

We wish to determine if the implementation of a program has been beneficial. Say we can use a count variable **M** to measure this aspect. For instance, **M** might be the yearly count of some very specific type of safety-related incidents that have not been numerically tracked in the past.

By evaluating any change in **M**, we hope to determine if there was a statistically measurable effect resulting from the program implementation. To do this, we also need a second count variable, **C**, to act as a "quasi-control"[14] or "gauge" in the sense that

a) The values of **C** are not affected by the program.
b) There are good, logical reasons for believing that, except for the effects of the program, the variable **M** has the same pattern of (non-random) variation as does the variable **C**.
c) Except for any effect of the program on the values of **M**, the values of **C** and **M** are reasonably proportional.

If a variable **C** is identified that we believe will satisfy these conditions, we then conduct the focused study to obtain the data we need, namely numerical values

[14] If the project is not implemented NAS-wide, one possibility for a "control" or "predictor" is "before" and "after" metric values for areas not impacted. The "metric of interest" values should then be for only those areas impacted by the project.

for the variables **M** and **C** in periods both before implementation of the program and after full implementation of the program.

For example, we might review, week-by-week, written incident reports, in each case determining if the report is an instance of **M**, an instance of **C**, or neither. The value of $\mathbf{M_b}$ would then be the number of "before program implementation" reports that included the condition measured by **M.** The other values, $\mathbf{M_a}$, $\mathbf{C_b}$, and $\mathbf{C_a}$ would be similarly obtained. (We also would keep a record of the week-by-week paired $\mathbf{(M_b,C_b)}$ values for a test of the viability of **C** as a "quasi-control" for **M**.)

We now describe how the data are used to estimate the impact of the program.

We construct a 2×2 table of the count data, as below.

	M	**C**
After program implemented	$\mathbf{M_a}$	$\mathbf{C_a}$
Before program initiation	$\mathbf{M_b}$	$\mathbf{C_b}$

The idea is quite simple (although it took two months to think of it). If the program has no impact, we would expect the values of **M** and **C** before the program's initiation to be proportional to the values after the program was fully operational. That is,

$$\frac{\mathbf{M_a}}{\mathbf{M_b}} = \frac{\mathbf{C_a}}{\mathbf{C_b}} \quad .$$

If program did have an impact, then the post implementation values of **M** should change relative to **C**. That is, there should be a change in the value of $\mathbf{M_a}$ relative to $\mathbf{C_a}$. For example, if **M** represents the number of safety incidents, and the program had a positive impact on the occurrence of these incidents, then $\mathbf{M_a}$ would be smaller than it would be if the program had no impact. That is, the proportion above would no longer hold, and we would have

$$\frac{M_a}{M_b} < \frac{C_a}{C_b}$$

If we then "add back" $\boldsymbol{M_x}$, the "invisible" numerical impact of the program on **M**, we restore the proportion:

$$\frac{M_a + \boldsymbol{M_x}}{M_b} = \frac{C_a}{C_b}.$$

$\boldsymbol{M_x}$ is the number of incidents that "didn't happen" during the period studied because of the beneficial impact of the program.

We can illustrate this by a table,

	M	**C**
After implementation	$\boldsymbol{M_x}$ / $\mathbf{M_a}$	$\mathbf{C_a}$
Before implementation	$\mathbf{M_b}$	$\mathbf{C_b}$

We call this table an I*mpact Assessment Diagram.*

Note that if the program improvement <u>increased</u> the value of M, then the post-implementation value of **M** will be larger than would be the case without the program. Therefore, the value $\boldsymbol{M_x}$ will be <u>negative</u>.

$\boldsymbol{M_x}$ is an "unseen" number, but we can solve the above proportion for it.

$$\boldsymbol{M_x} = \frac{\mathbf{M_b} \times \mathbf{C_a}}{\mathbf{C_b}} - \mathbf{M_a} = \frac{\mathbf{M_b} \times \mathbf{C_a} - \mathbf{M_a} \times \mathbf{C_b}}{\mathbf{C_b}} \quad .$$

$\boldsymbol{M_x}$, is a numerical count and only has meaning relative to the value $\mathbf{M_a}$. A more useful number is the effectiveness $\boldsymbol{e}$ of the program, expressed as the relative fraction of improvement due to the program,

$$\boldsymbol{e} = \frac{\text{actual improvement as a result of the program}}{\text{what would have occurred without the program}}$$

$$= \left| \frac{\boldsymbol{M_x}}{\boldsymbol{M_x} + \mathbf{M_a}} \right|$$

$$= \left| \frac{\mathbf{M_b C_a} - \mathbf{M_a C_b}}{\mathbf{M_b C_a}} \right| \quad .$$

The absolute value is used to ensure that the value of $\boldsymbol{e}$ is always positive. (Otherwise, it would be negative if improvement meant increased values of **M**.)

Until now, we have not made a clear distinction between *population* and *sample*. If we could collect <u>all</u> of the data on **M** and **C** for all time, we would have the *population* data. But practically, we only collect data for the periods covered by our focused study, that is, we only have *sample* data. So, we cannot determine the true value of the actual effectiveness $\boldsymbol{e}$ of the program, for that would require our having all of the data. Instead, we have to be satisfied with obtaining an estimate of the actual effectiveness $\boldsymbol{e}$ by using the *sample* data from our study.

In the remainder of this discussion, the variables **M** and **C** will refer to *sample* data, and we will use the symbol $\hat{\boldsymbol{e}}$ to denote a sample data estimate of the actual effectiveness $\boldsymbol{e}$. The equation for calculating $\hat{\boldsymbol{e}}$ is the same as for $\boldsymbol{e}$, except that the symbols **M** and **C** refer to *sample* data.

That is,

$$\hat{\boldsymbol{e}} = \left| \frac{\boldsymbol{M}_x}{\boldsymbol{M}_x + \mathbf{M_a}} \right|$$

$$= \left| \frac{\mathbf{M_b C_a - M_a C_b}}{\mathbf{M_b C_a}} \right| \quad ,$$

where all of the **M** and **C** data are sample values.

For example, if our study yielded the following data,

	M	C
After implementation	**6**	**22**
Before implementation	**14**	**36**

then,

$$\hat{\boldsymbol{e}} = \left| \frac{14\times22-6\times36}{14\times22} \right| = 0.2987 \approx 0.30 = 30\% \quad .$$

These data suggest that the program improved the values of **M** by an estimated 30%.

But, are the data sufficient to show that the program actually is beneficial? That is, to show that the actual (i.e., population) effectiveness is positive. Is $\boldsymbol{e} > \mathbf{0}$? To do this, we must first show that the variable **C** is a good quasi-control variable for **M,** and then that the value of $\boldsymbol{e}$ is statistically significant.

Step 1: <u>Be sure all of the conditions listed for the test in Table 1 are satisfied and that your data are not moving average values.</u>

Step 2: Hopefully, you can find a variable (metric) **C** that can serve as a "quasi-control" variable for the variable **M**.[15] If you can't find such a variable, stop—this test cannot be used.

Step 3: You now begin the focused study and record the numerical counts of **M** and **C** covering a time period before the initiation of the program and after its full implementation. You should divide the pre-implementation time period into several parts, collecting the counts of **M** and **C** for each so that you can test **C** in Step 4 to determine if it is a reasonable control variable. (To save resources, you might wish to postpone the study of a post-implementation period until after you have performed the test in Step 4. If you do so, do not let more than a couple of days pass before you resume the study and use the same staff to conduct both parts, so results are consistent.)

Step 4: The limited data available that caused you to choose this test also usually precludes any good test for the suitability of **C** as a "quasi-control" variable. However, if you can separate the pre-implementation data into several distinct, per-period $(\mathbf{C_i},\mathbf{M_i})$ pairs, use the procedure in Test D, Step 4. If that test is passed, proceed to Step 5, below. If you have insufficient data to use the procedure in Test D, Step 4, the following procedure will provide some assurance that **C** is not a "terrible" choice for a "quasi-control."

a) Divide the pre-implementation period in half and enter the counts for **M** and **C** in each half-period in the table below. Also enter the row sums, column sums and grand total in this table.

	M	**C**	**Totals**
First half of pre-implementation period	$\mathbf{M_1}$	$\mathbf{C_1}$	$\mathbf{T_1}$
Second half of pre-implementation period	$\mathbf{M_2}$	$\mathbf{C_2}$	$\mathbf{T_2}$
Totals	$\mathbf{T_M}$	$\mathbf{T_C}$	**N**

Note that $\mathbf{N = T_M + T_C = T_1 + T_2}$.

b) Calculate each of the following values:

15 If the project is not implemented NAS-wide, one possibility for a "control" or "predictor" is "before" and "after" metric values for areas not impacted. The "metric of interest" values should then be for only those areas impacted by the project.

$$\frac{T_M T_1}{N}\ ,\quad \frac{T_M T_2}{N}\ ,\quad \frac{T_C T_1}{N}\ ,\quad \frac{T_C T_2}{N}\ .$$

Each of these calculated values must be $\geq$ **5 .**

If they are not, either collect more data or use Fisher's Exact Test. (For Fisher's Exact Test you will need appropriate statistical software, and you will use the values $\mathbf{M_1}$, $\mathbf{M_2}$, $\mathbf{C_1}$, $\mathbf{C_2}$. Fisher's exact test will give you a significance level, so you should skip Step 4(c) and go to Step 4(d).)

c) Calculate the following (chi-square) value[16]

$$\chi^2 = \frac{N\left(M_1C_2 - M_2C_1\right)^2}{T_M \times T_C \times T_1 \times T_2}\ .$$

d) If $\chi^2 \leq \mathbf{0.7}$ [17] (or Fisher's Exact Test produced a significance level $\geq$ **0.4**) and there are good, logical reasons for believing that, except for the effects of the program, the variable **C** has the same pattern of (non-random) variation as does the variable **M**, then **C** is probably a reasonable choice for a "quasi-control" variable. If these conditions are not met, then **C** may not be appropriate for use. It's probably time to find a good statistician.

Step 5: If **C** satisfies the test in Step 4 (or in Test D, Step 4), it can now be used as a "quasi-control" in a test to determine if the project had a statistically significant impact on the variable **M**.

Step 5 will determine if we can say that the program has a beneficial impact. It will determine if the value of the actual effectiveness $\boldsymbol{e}$ is statistically significance, that is, that $\boldsymbol{e} > \mathbf{0}$.

This test is in multiple parts.

a) Using a table of the pre-implementation and post-implementation values you obtained in the focused study, add row and column totals.

[16] The "continuity correction" has been purposely omitted, as evidence suggests it gives poorer results.
[17] This is an ad hoc "test." To the best of the author's knowledge, there is no formal test to "prove" homogeneity (as opposed to "proving" nonhomogeneity).

	M	**C**	**Totals**
After implementation	$\mathbf{M_a}$	$\mathbf{C_a}$	$\mathbf{T_a}$
Before implementation	$\mathbf{M_b}$	$\mathbf{C_b}$	$\mathbf{T_b}$
Totals	$\mathbf{T_M}$	$\mathbf{T_C}$	**N**

Note that $N = T_a + T_b = T_M + T_C$.

b) Calculate each of the following values:

$$\frac{T_M T_a}{N}, \quad \frac{T_M T_b}{N}, \quad \frac{T_C T_a}{N}, \quad \frac{T_C T_b}{N} \quad .$$

Each of these calculated values must be $\geq$ **5 .**

If they are not, either collect more data or use Fisher's Exact Test. (For Fisher's Exact Test you will need appropriate statistical software, and you will use the values $\mathbf{M_a}$, $\mathbf{M_b}$, $\mathbf{C_a}$, $\mathbf{C_b}$. Fisher's exact test will give you a significance level, so you will not perform Step 5(c), but will proceed to Step 6.)

c) Calculate the following (chi-square) value[18]

$$\chi^2 = \frac{N\left(M_a C_b - M_b C_a\right)^2}{T_M \times T_C \times T_a \times T_b} \quad .$$

Calculate $\mathbf{z} = -\sqrt{\chi^2}$.

In Table F-1, find the significance level corresponding to the value **z.**

Step 6: If the significance level is 10% or better, you can state that, "this test indicates, at the ____% significance level, that the project has had a beneficial impact and improved [the situation] by an estimated factor of $\hat{\varepsilon}$**,"** where

$$\hat{\varepsilon} = \left| \frac{M_b C_a - M_a C_b}{M_b C_a} \right| \quad .$$

Step 7: Note that in Step 6, although our estimated effectiveness, $\hat{\varepsilon}$, may be a large value, we have only "proven" that the actual effectiveness $\boldsymbol{\varepsilon} > \mathbf{0}$.

[18] The continuity correction has been purposely omitted, as evidence suggests it gives poorer results

At the beginning of this appendix, in the paragraph just above the heading The Statistical Tests**,** we said, "if you wish to statistically determine whether there has been a practical change, you must first decide on how big the change has to be in order to be considered practical. That is, you must select a minimum value that must be achieved in order to say a change was practical. Then, a statistical significance test is used to determine if the post-change data are so different that it is very unlikely that the change was less than the minimum you selected. Statistical tests for a *practical* change are somewhat more complex than tests for *some* change. If you wish to test for a practical change, we suggest that you employ a statistician."

Because this test (Test E) cannot be found in statistics books, we present here the steps for testing for "practical significance."

a) Do not proceed unless Step 5, above, yielded a significance level of 10% or better and you were not required to use Fisher's exact test.[19]

b) Determine the smallest value of $\boldsymbol{e}$ that would be considered to be of practical significance. Call this value $\boldsymbol{e}_{\mathbf{o}}$.

c) Calculate the value

$$z = \frac{-\ln\left(\dfrac{C_a M_b}{M_a C_b}\left[1-\boldsymbol{e}_o\right]\right)}{\sqrt{\dfrac{1}{M_a}+\dfrac{1}{M_b}+\dfrac{1}{C_a}+\dfrac{1}{C_b}}} .$$ [20]

d) Compare this value of **z** with the values in Table F-1 to obtain the statistical significance level. If the significance level is 10% or better, you can state that, "a __% significance level statistical test of the data indicate that the program has a beneficial impact with an effectiveness of at least $\boldsymbol{e}_{\mathbf{o}}$, and with an estimated actual effectiveness of $\hat{\boldsymbol{e}}$," where

$$\hat{\boldsymbol{e}} = \left|\frac{M_b C_a - M_a C_b}{M_b C_a}\right| .$$

[19] If you were required to use Fisher's Exact test, you do not have enough data for the equation in Step 8(c) to be accurate.

[20] **Warning!** If you are using Excel®, be aware that the definitions of the log function in Excel and in Excel Visual Basic® differ.

Example: Suppose that the FAA has a program that should reduce some, but not all occurrences of a particular type of incident. In particular, the program is expected to have substantially reduced some types of pilot error that cause this type of incident, but will not have affected other types of pilot error that cause this type of incident.

Data has been tracked on this type of incident, but not on the types of pilot error that cause it. However, this information is available in the reports of the incidents. A focused, two-month, pre-implementation study of these incident reports is conducted, and a one-month post-implementation study is also conducted. In this study, each incident report is reviewed and is classified as either an incident that should have been eliminated by the program or should not have been affected by the program.

The monthly counts of the "should have" and the "should not have" incidents are tabulated in the table below.

	"Should Have" **M**	"Should Not Have" **C**
Pre-implementation, Month 1	$M_{b1} = 21$	$C_{b1} = 34$
Pre-implementation, Month 2	$M_{b2} = 17$	$C_{b2} = 28$
Post-implementation	$M_a = 6$	$C_a = 22$

We believe that the right, "**C**" column data values can serve as a "quasi-control" so we proceed with Step 4 to help give some assurance to this belief.

Step 4(a): We first analyze the pre-implementation data to determine if the variable **C** can serve as a "quasi control" variable. We add to the table the row and column sums.

	"Should Have" **M**	"Should Not Have" **C**	Totals
Pre-implementation, Month 1	$M_{b1} = 21$	$C_{b1} = 34$	$T_1 = 55$
Pre-implementation, Month 2	$M_{b2} = 17$	$C_{b2} = 28$	$T_2 = 45$
Totals	$T_M = 38$	$T_C = 62$	$N = 100$

Step 4(b): We calculate

$$\frac{T_M T_1}{N} = \frac{38 \times 55}{100} = 20.9$$

$$\frac{T_M T_2}{N} = \frac{38 \times 45}{100} = 17.1$$

$$\frac{T_C T_1}{N} = \frac{62 \times 55}{100} = 34.1$$

$$\frac{T_C T_2}{N} = \frac{62 \times 45}{100} = 27.9 \quad .$$

All of these values are greater than 5, so the condition in Step 4(b) is satisfied and we proceed to Step 4(c) to calculate the c^2 value.

Step 4(c)

$$c^2 = \frac{N\left(M_1 C_2 - M_2 C_1\right)^2}{T_M \times T_C \times T_1 \times T_2}$$

$$= \frac{100\left(\left(21 \times 28\right) - \left(17 \times 34\right)\right)^2}{38 \times 62 \times 55 \times 45}$$

$$= \mathbf{0.0017}$$

Step 4(d): $c^2 \leq 0.7$ and we have good, logical reasons for believing that, except for the effects of the program, the variable **C** has the same pattern of (non-random) variation as does the variable **M**. So we are willing to use **C** as a "quasi-control" variable, and we proceed to Step 5..

Step 5: In this multi-part step we determine, by use of a statistical significance test, if we can say that the program has a beneficial impact. That is, if the data and test indicate that the actual effectiveness $\boldsymbol{e} > \mathbf{0}$.

Step 5 (a): We first add row and column totals to the table of post- and pre-implementation data values.

	"Should Have" **M**	"Should Not Have" **C**	Totals
Post-implementation	$M_a = 6$	$C_a = 22$	$T_a = 28$
Pre-implementation	$M_b = 38$	$C_b = 62$	$T_b = 100$
Totals	$T_M = 44$	$T_C = 88$	$N = 128$

<u>Step 5 (b)</u>: We now calculate

$$\frac{T_M T_a}{N} = \mathbf{\frac{44 \times 28}{128} = 9.625}$$

$$\frac{T_M T_b}{N} = \mathbf{\frac{44 \times 100}{128} = 34.375} \quad ,$$

$$\frac{T_C T_a}{N} = \mathbf{\frac{88 \times 28}{128} = 19.25} \quad ,$$

$$\frac{T_C T_b}{N} = \mathbf{\frac{88 \times 100}{128} = 68.75} \quad .$$

All of these values are ≥ 5, so we can proceed to Step 5(c).

<u>Step 5 (c)</u>: We now calculate the chi-square value,

$$c^2 = \frac{N\left(M_a C_b - M_b C_a\right)^2}{T_M \times T_C \times T_a \times T_b}$$

$$= \frac{128\left(6 \times 62 - 38 \times 22\right)^2}{44 \times 88 \times 28 \times 100}$$

$$= \mathbf{2.542} \ ,$$

calculate

$$\mathbf{z} = -\sqrt{c^2} = -\sqrt{2.542} = -1.59 \, ,$$

and compare this value with those in Table F-1. Table F-1 indicates that we have a significance level of better than 10%, so we proceed to Step 6.

Step 6: We can now make a statement about the effectiveness of the program. First we calculate

$$\hat{e} = \left| \frac{M_x}{M_x + \mathbf{M_a}} \right|$$

$$= \left| \frac{M_b C_a - M_a C_b}{M_b C_a} \right|$$

$$= \left| \frac{38 \times 22 - 6 \times 62}{38 \times 22} \right|$$

$$= 0.555 \ .$$

Now, we can state that, "this test indicates, at a better than 10% significance level, that the project has had a beneficial impact and has reduced the targeted, pilot errors and their resulting incidents by an estimated 55.5% ."

Step 7: While our estimate of the project's effectiveness, based on our sample of data, is $\hat{e}$ = 55.5%, the significance test results in Step 6 only "proved" that the actual effectiveness is positive, that is, that $e > 0$.

In this project's Investment Analysis, it was estimated that the project would have to achieve a minimum benefit of $\mathbf{B_o}$ for the project to break even. The value $\mathbf{B_o}$ expressed as an effectiveness index value, e_o, is $e_o = 20\%$.

We now determine if we can say, with statistical significance, that the actual project effectiveness $e \geq 20\%$.

The statistical significance level (for $e > 0$) found in Step 6 was better than 10% and we did not use Fisher's exact test, so we can proceed with calculating

$$z = \frac{-\ln\left(\frac{C_a M_b}{M_a C_b}[1-\boldsymbol{e_o}]\right)}{\sqrt{\frac{1}{M_a}+\frac{1}{M_b}+\frac{1}{C_a}+\frac{1}{C_b}}} \quad .$$

$$= \frac{-\ln\left(\frac{22\times 38}{6\times 62}[1\text{-}0.2]\right)}{\sqrt{\frac{1}{6}+\frac{1}{38}+\frac{1}{22}+\frac{1}{62}}}$$

$$= \mathbf{-1.16}\ .$$

If we compare this value with those in Table F-1, we see that the significance level is not even 10%.

So, although the project has had some benefit and although our estimated effectiveness is $\hat{\boldsymbol{e}} = 55.5\%$, we have not been able to show that the actual effectiveness $\boldsymbol{e} \geq \mathbf{20\%}$.

Test F: Box-Jenkins-Tiao Intervention Analysis with possible multivariate transfer function components.

This is a highly sophisticated statistical procedure that is quite adaptable and powerful. It should only be performed by someone thoroughly familiar with using it.[21]

Test X: One-sided, large sample test for a mean value.

This test should only be used when there are very good reasons for believing that if the project had not been implemented, the metric values would have become significantly worse than would have been predicted by extrapolating pre-implementation metric values. The test compares the average of post-implementation metric values **m** with the estimated average of what the values would have been without the project's implementation.

Step 1: <u>Be sure all of the conditions listed for the test in Table 1 are satisfied and that your data are not moving average values.</u>

[21] Even then, mistakes are sometimes made. See: Cohen, S. "A Common Error In Time Series Intervention Analysis," *AIIE (American Institute of Industrial Engineers) Transactions*, vol. 14, no. 2, June 1982.)

Step 2: Record at least 30 period-by-period (e.g., monthly) metric values after the project is fully operational. Denote these values by

$$\mathbf{m_1, m_2, m_3, \ldots, m_n} \ .$$

n is the number of metric values recorded.

Step 3: In the Benefit Analysis portion of the Investment Analysis, the reference case estimates included a prediction of the values the metric **m** would have if the project were not implemented. Use these predictions to estimate what the reference case, non-implementation values of the metric **m** would have been for the same periods as chosen in Step 2. Denote the average of these, non-implementation values by $\mathbf{m_0}$.

Step 4: Calculate the following values.

$$\bar{\mathbf{m}} = \frac{\sum_{i=1}^{n} \mathbf{m_i}}{\mathbf{n}}$$

$$\mathbf{s} = \sqrt{\frac{\sum_{i=1}^{n} \left(\mathbf{m_i} - \bar{\mathbf{m}}\right)^2}{\mathbf{n - 1}}}$$

$\bar{\mathbf{m}}$ is the average (mean) of the **n** post-implementation metric values.
s is the estimated (sample) standard deviation of the post-implementation metric values.

Step 5: Compare $\bar{\mathbf{m}}$ and $\mathbf{m_0}$.

a) If larger metric values indicate improvement and if $\bar{\mathbf{m}} < \mathbf{m_0}$, then based on the original benefit analysis, the project may have made things worse. Stop. You cannot claim a positive benefit.

If $\bar{\mathbf{m}} > \mathbf{m_0}$, proceed to Step 6.

b) If smaller metric values indicate improvement and if $\bar{\mathbf{m}} > \mathbf{m_0}$, then, based on the original benefit analysis, the project may have made things worse. Stop. You cannot claim a positive benefit.

If $\bar{\mathbf{m}} < \mathbf{m_0}$, proceed to Step 6.

Step 6: a) Calculate

$$\mathbf{z} = \frac{-|\bar{\mathbf{m}} - \mathbf{m}_0|}{\left(\frac{\mathbf{s}}{\sqrt{\mathbf{n}}}\right)} \quad .$$

b) Compare this value of $\mathbf{z}$ with the values in Table F-1 to obtain the statistical significance level.

c) If the test yields a significance level of at least 10%, you can state that, "this test indicates that at the ____% significance level, the project has had a beneficial impact and an estimate of the average per-period benefit is the value $|\bar{\mathbf{m}} - \mathbf{m}_0|$."

(this page intentionally blank)

Table F-1
One-Sided Test Significance Levels for Larger Samples
(based on the Normal [Gaussian] distribution)

	Significance level *
z < - 3.09	**Better than 0.1%**
- 3.09 £ z < - 2.58	**Better than 0.5%**
- 2.58 £ z < - 2.33	**Better than 1.0%**
- 2.33 £ z < - 1.96	**Better than 2.5%**
- 1.96 £ z < - 1.65	**Better than 5.0%**
- 1.65 £ z < - 1.28	**Better than 10%**
z ≥ - 1.28	**Worse than 10%**

* ”Better than” means “less than.” Example: 5% is better than 10%.

Table F-2
Critical Values of T in the Wilcoxon
Matched-Pairs Signed-Ranks Test

N	**Level of significance for one-tailed test**				
	10%	**5%**	**2.5%**	**1%**	**0.5%**
6	4	2	0	—	—
7	6	4	2	0	—
8	8	6	4	2	0
9	11	8	6	3	2
10	14	11	8	5	3
11	18	14	11	7	5
12	22	17	14	10	7
13	26	21	17	12	10
14	31	26	21	16	13
15	37	30	25	19	16
16	42	36	30	23	20
17	49	41	35	28	23
18	55	47	40	33	28
19	62	53	46	38	32
20	70	60	52	43	37
21	77	68	59	49	43
22	86	75	66	56	49
23	95	83	73	62	55
24	104	92	81	69	61
25	114	101	89	77	68
26	124	110	137	114	76
27	135	120	107	93	83
28	146	130	117	102	92
29	157	141	127	111	100
30	169	152	137	120	109

Based on Table G in Siegel, S., *Nonparametric Statistics for the Behavioral Sciences*, New York: McGraw-Hill, 1956; which was adapted from Table I of Wilcoxon F., *Some rapid approximate statistical procedures*. New York: American Cyanamid Company, 1949, p.13. Addition of rows 26-30 by Stephen Cohen, FAA, September 2000.

Table F-3

Significance Levels for the One-Sided Sign Test

N \ C	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
5	.031	.188	.500	.812	.969	≈1.0										
6	.016	.109	.344	.656	.891	.984	≈1.0									
7	.008	.062	.227	.500	.773	.938	.992	≈1.0								
8	.004.	.035	.145	.363	.637	.855	.965	.996	≈1.0							
9	.002	.020	.090	.254	.500	.746	.910	.980	.998	≈1.0						
10	.001	.011	.055	.172	.377	.623	.828	.945	.989	.999	≈1.0					
11		.006	.033	.113	.274	.500	.726	.887	.967	.994	≈1.0	≈1.0				
12		.003	.019	.073	.194	.387	.613	.806	.927	.981	.997	≈1.0	≈1.0			
13		.002	.011	.046	.133	.291	.500	.709	.867	.954	.989	.998	≈1.0	≈1.0		
14		.001	.006	.029	.090	.212	.395	.605	.788	.910	.971	.994	.999	≈1.0	≈1.0	
15			.004	.018	.059	.151	.304	.500	.696	.849	.941	.982	.996	≈1.0	≈1.0	≈1.0
16			.002	.011	.038	.105	.227	.402	.598	.773	.895	.962	.989	.998	≈1.0	≈1.0
17			.001	.006	.025	.072	.166	.315	.500	.685	.834	.928	.975	.994	.999	≈1.0
18			.001	.004	.015	.048	.119	.240	.407	.593	.760	.881	.952	.985	.996	.999
19				.002	.010	.032	.084	.180	.324	.500	.676	.820	.916	.968	.990	.998
20	.			.001	.006	.021	.058	.132	.252	.412	.588	.748	.868	.942	.979	.994
21				.001	.004	.013	.039	.095	.192	.332	.500	.668	.808	.905	.961	.987
22					.002	.008	.026	.067	.143	.262	.416	.584	.738	.857	.933	.974
23					.001	.005	.017	.047	.105	.202	.339	.500	.661	.798	.895	.953
24					.001	.003	.011	.032	.076	.154	.271	.419	.581	.729	.846	.924
25						.002	.007	.022	.054	.115	.212	.345	.500	.655	.788	.885
26						.001	.005	.014	.038	.084	.163	.279	.423	.577	.721	.837
27						.001	.003	.010	.026	.061	.124	.221	.351	.500	.649	.779
28							.002	.006	.018	.044	.092	.172	.286	.425	.575	.714
29							.001	.004	.012	.031	.068	.132	.229	.356	.500	.644
30							.001	.003	.008	.021	.049	.100	.181	.292	.428	.572

Reproduction of Table D in Siegel, S., *Nonparametric Statistics for the Behavioral Sciences*, New York: McGraw-Hill, 1956; which was adapted from Table IV-B of Walker, Helen and Lev, J., *Statistical* Inference. New York: Holt, 1953, p.458. Addition of rows 26-30 by Stephen Cohen, FAA, September 2000.

(this page intentionally blank)

APPENDIX G
REFERENCES

Architecture and System Engineering Directorate (ASD-100), *Capability Architecture Tool Suite ¾ Intranet (CATS-I)*, http://172.27.164.125/CATS/CATSI.cfm , (frequently updated). This "home" page leads to several useful references.

Architecture and System Engineering Directorate (ASD-100), *National Airspace System Architecture Version 4.0*, Washington: Federal Aviation Administration, January 1999. [http://172.27.164.125/CATS/Tutorials/NASArch.htm .]

Bolczak, C.N., L.M. Brown, J.H. Hoffman, E.S. Lacher, *System Performance Baseline: Initial Indicators, Analyses, and Results*, Report WN 96-W0000015, McLean, VA: MITRE CAASD, June 1996.

Citrenbaum, D. and R. Juliano, *A Simplified Approach to Baselining Delays and Delay Costs for the National Airspace System (NAS)*, Interim Report 12a (DCN-R80406-02), Washington: Federal Aviation Administration, August 1998 – updated May 1999.

CNS/ATM Focused Team (C/AFT), *Airline Metric Concepts For Evaluating Air Traffic Service Performance*, Report of the Air Traffic Services Performance Focus Group (ATSP FG), February 1, 1999.

Cohen, S. "A Common Error In Time Series Intervention Analysis," *AIIE (American Institute of Industrial Engineers) Transactions*, vol. 14, no. 2, June 1982.)

Dorfman, G., W,K. MacReynolds, F.R. Morser, E.J. Spear, *Evaluating Benefits of a Change in NAS Performance: How to Include Cost and Revenue Implications*, (MP 00W0000156)McLean, VA: MITRE CAASD, September 1999.

FAA Past Performance Database, Washington: Federal Aviation Administration, http://www.faa.gov/pastperf .

Federal Aviation Administration Acquisition System Toolset (FAST), Washington: FAA Intranet, http://fast.faa.gov , (updated monthly).

FFP1, *Performance Metrics: An Operational Impact Evaluation Plan, Version 1.0*, Washington: Free Flight Phase 1 (FAA and RTCA), August 12, 1999.

FFP1, *Performance Metrics ¾ Results to Date, June 2000 Report*, , Washington: Free Flight Phase 1 (FAA and RTCA), June 2000.

Office of Aviation Policy and Plans, *Economic Analysis of Investment and Regulatory Decisions ¾ Revised Guide*, FAA-APO-98-4, Washington: Federal Aviation Administration, January 1998.

Office of Aviation Policy and Plans, *Economic Values for Evaluation of Federal Aviation Administration Investment and Regulatory Programs*, FAA-APO-98-8, Washington: Federal Aviation Administration, June 1998. [The latest version (as of May 2000) of this guide, which includes an additional chapter not present in the paper version, may be found at http://api.hq.faa.gov/apo_pubs.htm#ANCHOR98_10 .]

Office of Management and Budget, *OMB Circular A-94*, Washington: Office of the President. [http://www.whitehouse.gov/OMB/circulars/a094/a094.html].

Operations Assessment Division, DTS-59, *Cost, Benefit, and Risk Assessment Guidelines for R,E&D Investment Portfolio Development*, Report No. WP-43-FA92F-99-1, Cambridge: Volpe National Transportation Systems Center, October 1998.

Operations Assessment Division, DTS-59, *Risk Assessment Guidelines for the Investment Analysis Process ¾ Update of July 1999*, Report No. WP-59-FA7N1-97-2, Cambridge: Volpe National Transportation Systems Center, July 1999.

Operations Assessment Division, DTS-43, *Sample Benefit and Cost Assessment with Explanations ¾ Companion Document to "Cost, Benefit, and Risk Assessment Guidelines for R,E&D Investment Portfolio Development," November 1997*, Report No. WP-43-FA82F-98-2, Cambridge: Volpe National Transportation Systems Center, November 1977.

Osborne, Tony, *Department Of Transportation/Federal Aviation Administration Acquisition And Program Risk Management Guidance*, FAA-P-l8l0, Revision 2, Washington: Federal Aviation Administration, December 20, l996.

Siegel, S., *Nonparametric Statistics for the Behavioral Sciences*, New York: McGraw-Hill, 1956.

==

Web Sites for Other Useful Information

Aviation Glossary : http://172.27.164.125/CATS/Search/default.cfm?SG=TRUE

FAA Architecture home page: http://www.nas-architecture.faa.gov .

This page has links to several pages including the must-see *Capability Architecture Tool Suite (CATS)*. Note that the version of CATS accessible from the home page may be different from the private FAA page, http://172.27.164.125/cats/

The *FAA National Aviation Research Plan* (formerly the *RE&D Plan*): http://172.27.164.125/CATS/Tutorials/NARP.htm

The *NAS Blueprint*: http://172.27.164.125/CATS/Tutorials/Blueprint.htm

Other Architecture-related documents: http://172.27.164.125/CATS/Tutorials/Other-Intro.htm

Useful APO publications, data bases, and information may be found at
http://api.hq.faa.gov/apo_pubs.htm and at http://www.apo.data.faa.gov/

OMB guidance circulars: http://www.whitehouse.gov/OMB/circulars/index.html